

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : <b>C12N 15/12, C07K 14/47, 14/475, G01N 33/68, C12Q 1/68, A61K 31/70, 38/17</b>		A2	(11) International Publication Number: <b>WO 99/37764</b> (43) International Publication Date: <b>29 July 1999 (29.07.99)</b>
(21) International Application Number: <b>PCT/EP99/00329</b>		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AR IPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: <b>20 January 1999 (20.01.99)</b>		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(30) Priority Data: <b>98200226.3 27 January 1998 (27.01.98) EP</b>			
(71) Applicant ( <i>for all designated States except US</i> ): VLAAMS INTERUNIVERSITAIR INSTITUUT VOOR BIOTECHNOLOGIE [BE/BE]; Rijwijschestraat 118, Bus 1, B-9052 Zwijnaarde (BE).			
(72) Inventors; and (75) Inventors/Applicants ( <i>for US only</i> ): VBUGELERS, Mark, Paul, Dittmar [BE/BE]; Schreverland 22, B-3530 Heusden-Zolder (BE). DAVID, Guido, Joseph, Frans [BE/BE]; Koningin Elizabethlaan 11, B-3000 Leuven (BE).			
(74) Agent: DE CLERCQ, Ann; Ann De Clercq & Co. B.V.B.A., Brandstraat 100, B-9830 Sint-Martens-Latem (BE).			
(54) Title: NEW MEMBERS OF THE GLYCAN GENE FAMILY			
(57) Abstract			
<p>The invention relates to a novel polynucleotide encoding a new glycan-related protein ("glycan-6") and the gene for glycan-4 as well as derivatives of both genes for use in methods of diagnosis and therapy. Derivatives comprise for example fragments of the gene either isolated or synthetic and having a length that is smaller than the complete gene; primers, comprising at least 10 consecutive gene specific nucleotides, preferably about 20 gene specific consecutive nucleotides of the nucleotide sequence of the gene; longer oligonucleotides up to the full length of the gene; antisense variants of the gene, the fragments or the primers; antibodies directed to the gene, fragments, primers or complementary strands thereof; any specific ligand for DNA that can be used as a specific probe, peptide nucleic acid probes.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroun	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## NEW MEMBERS OF THE GLYPLICAN GENE FAMILY

The present invention relates to the characterization and chromosomal localization of new members of the glypcan gene family and to the use of members of this family in diagnostics and/or 5 therapeutics.

Glypicans are glypiated cell surface heparan sulfate proteoglycans, the first of which was originally identified in human lung fibroblasts (David et al., 1990). The known five members of this family have similar 10 core protein sizes (about 60 kDa), share a unique and very conserved cysteine spacing, and are linked to the cell membrane by a glycosyl phosphatidyl inositol (GPI)- anchor. Those five known glypicans are of vertebrate origin and include glypcan (glypcan-1, David et al., 15 supra), Cerebroglycan (glypcan-2, Stipp et al., 1994), OCI-5 (glypcan-3, Filmus et al., 1995), K-glypcan (glypcan-4, Watanabe et al., 1995) and glypcan-5 (Veugelers et al., 1997).

All the structural features of the vertebrate 20 glypicans are also present in the product of dally (division abnormally delayed), a locus identified in Drosophila melanogaster by genetic screening for mutants affecting cell division patterning in the developing central nervous system (Nakato et al., 1995). Besides 25 disturbing cell cycling in the nervous system dally mutations also affect viability and produce morphological defects in several adult tissues, including the eyes, antennae, wings and genitalia. The dally mutants, the well established co-receptor activities of the cell 30 surface proteoglycans for various ligands that are known to mediate developmental instructions, and the tissue and stage-specific expressions of the glypicans, all implicate the glypcan group of integral membrane proteoglycans in the control of cell division and 35 patterning during development. This contention has recently been corroborated by the identification of mutations in GPC3, the gene coding for the human

homologue of OCI-5 (glypican-3), that cause the Simpson-Golabi-Behmel overgrowth syndrome (SGBS) (Pilia et al., 1996). This X-linked condition, which clinically has to be differentiated from the autosomal Beckwith-Wiedeman Syndrome, is characterized by pré- and post-natal overgrowth with visceral and skeletal anomalies, and is associated with a high risk for developing embryonal tumors, including Wilms' tumor and neuroblastoma.

It was therefore anticipated by the present inventors that chromosomal assignment of the genes for the members of the glypican family and the identification of potentially additional members in this family may be of general relevance for the understanding of somatic overgrowth and tumor predisposition. So far, only 15 glypican, the homologue of OCI-5 (glypican-3) and glypican-5 have been identified in human. The corresponding genes GPC1, GPC3 and GPC5 have been localized to chromosomes 2q35-q37, Xq26 and 13q32, respectively (Vermeesch et al., 1995; Pilia et al., 1996; 20 Veugelers et al., 1997). The cDNA nucleotide and derived amino acid sequences of these genes are given in figures 3, 4 and 5, respectively.

It is thus the object of the present invention to provide new members of the glypican family, and to 25 study their possible implications in various medical indications. It is a further object of the invention to use the information derivable from the members of the glypican gene family for designing diagnostic methods and kits and/or for the development of therapeutics.

30 In the research that led to the present invention two novel human cDNAs were identified encoding glypican-related proteins. The corresponding gene for the first was mapped to chromosome 13q32. In this application this gene will be identified as GPC6, whereas the protein 35 encoded by the gene will be called glypican-6. The predicted primary structure of the GPC6 protein was found to show significant sequence similarity to glypican (glypican-1), to the human homologue of OCI-5

(glypican-3), to glypican-5, to the glypican-related proteins Cerebroglycan (glypican-2 of the rat), to K-glypican (glypican-4 of the mouse) and to the gene product of the dally locus in Drosophila melanogaster.

5 The similarity pertains to a conserved sequence motif, present in all seven proteins and that include a set of 14 conserved cysteine residues found in specific positions. Additional similarities include the overall sizes of the proteins, the presence of N-terminal and

10 C-terminal signal peptide-like sequences (the first predicted to be involved in the membrane translocation of the nascent polypeptide, the second in the temporary membrane anchorage and subsequent glypiation of the proteins), and the presence of glycosaminoglycan

15 attachment consensus sequences close to the C-termini of the proteins.

Glypican-6 is, however, more similar to K-glypican and human glypican-4 (see below) than to the other glypicans. As to these other glypicans, glypican-6

20 is more similar to glypican-1 and glypican-2 than to glypican-3 and glypican-5. The gene encoding glypican-6 (GPC6) has a similar exon-intron organization as the gene encoding glypican-4 (GPC4 as was now found according to the invention) and the gene encoding glypican-1 (GPC1).

25 This organization differs from the exon-intron organization of the gene encoding glypican-3 (GPC3) and that of the gene encoding glypican-5 (GPC5), while GPC3 and GPC5, in turn, resemble one another in terms of their intron-exon organizations. This indicates the possible

30 existence of (at least two) glypican subfamilies: one comprising, so far, the glypicans 1, 2, 4 and 6; the other comprising, so far, the glypicans 3 and 5.

According to a further aspect of the invention, a cDNA is provided that encodes the human homologue of K-glypican (glypican-4), the corresponding gene of which localizes to chromosome Xq26 in very close proximity to the gene for glypican-3. Thus the GPC3 and GPC4 genes are adjacent, or near adjacent, to one another on chromosome

Xq26, while the GPC5 and GPC6 genes are adjacent, or near adjacent, to one another on chromosome 13q32. In these two examples a member (GPC3, respectively GPC5) of one of the glypcan subfamilies is physically linked to a member 5 (GPC4, respectively GPC6) of another glypcan subfamily. This indicates that the glypcan subfamilies and various members of these families may have arisen from the duplications of one ancestral glypcan gene and ancestral gene cluster. Maintenance of the physical associations 10 between these genes during evolution suggests that these genes may also be functionally linked.

Furthermore, it was found according to the invention that besides the gene for glypcan-3 (GPC3) also the gene for glypcan-4 (GPC4) may show aberrations 15 in patients having disorders and diseases involving abnormal cell growth and behavior, like somatic overgrowth and tumor formation. With this information various diagnostic tests could be developed in order to detect aberrations in the genes that encode glypcans and 20 aberrations in the expression levels of these genes. Moreover, this knowledge can be used to develop therapeutic compounds that restore the physical damage caused by the mutant gene.

The aberrations in the gene comprise for 25 example deletions or translocations within either or both of the two genes, but also mutations in either or both of them. These aberrations may lead to the absence of gene products or to abnormal gene products. Thus, the expression level of the gene may be used as another 30 parameter indicating the presence of one or more aberrant genes.

In the research that led to the invention various aberrations have been found in patients suffering from Simpson-Golabi-Behmel syndrome, including deletion 35 of the entire GPC4 gene and exons 7 and 8 of GPC3, deletion of exons 1 and 2 of GPC3, a T>A mutation in exon 3 of GPC3 leading to a W<sup>296</sup>>R substitution in glypcan 3, a C>T mutation in exon 8 of GPC4 leading to a A<sup>442</sup>>V

substitution in glycan 4, deletion of one dT nucleotide in GPC3 resulting in a frame shift mutation and premature termination of the protein.

The finding for GPC3 and GPC4 can be  
5 extrapolated to other members of the glycan gene family. GPC5 and GPC6 are likewise associated.

Aberrations in these genes can be identified in a similar manner as herein described for GPC3 and GPC4.

Various molecular biological techniques can be  
10 used to find these types of aberration in the genes.

These techniques include but are not limited to single strand conformation polymorphism (SSCP) screening, restriction fragment length polymorphism (RFLP) screening, gel electrophoresis, Southern blot analysis,  
15 PCR, DNA sequencing, etc.

Diagnostic methods according to the invention are based on the information derivable from the gene and/or its gene product. Such information comprises the nucleotide sequence, either sense or antisense, of the  
20 gene and the complementary strand thereof, and the amino acid sequence of the gene product encoded by the coding sequence of the gene. The information derivable from the gene or gene product can very well be defined by a person skilled in the art by referring to figures 1 to 6.

25 Figures 1 to 5 disclose the nucleotide sequence of the human cDNAs for glycans 1 and 3 to 6, as well as the derived amino acid sequence of the protein encoded by the cDNA. Figure 6 gives an alignment of the predicted amino acid sequences and the position of the exon boundaries  
30 for each of them. This information can be used to define so-called derivatives.

Derivatives of the nucleotide sequence of the gene are the gene itself, either isolated or synthetic, fragments of the gene, either isolated or synthetic and  
35 having a length that is smaller than the complete gene; primers, comprising at least 10 consecutive gene specific nucleotides, preferably about 20 gene specific consecutive nucleotides of the nucleotide sequence of the

gene; longer oligonucleotides up to the full length of the sequence of the gene; antisense variants of the gene, the fragments or the primers; antibodies directed to the gene, fragments, primers or complementary strands thereof; any specific ligand for DNA that can be used as a specific probe, peptide nucleic acid probes.

Other derivatives are transcripts (mRNA sequences) of the gene, from which in turn cDNA, antisense RNA, antisense cDNA, antibodies directed to the transcript, sense and antisense cDNA, antisense RNA and any specific ligand for RNA that can be used as a specific probe can be derived.

Derivatives of the amino acid sequence include the isolated or synthetic gene product (also called protein or polypeptide); isolated or synthetic peptides, comprising a specific sequence of consecutive amino acids encoded by the gene, antibodies directed to the gene product or peptides and any specific ligand for peptides that can be used as a specific probe.

Other derivatives are heparan sulfate chains or heparan sulfate structures, antibodies directed to heparan sulfate structures present on the product of the natural or synthetic gene as a result of the posttranslational modification of these gene products, any specific ligand for heparan sulfate that can be used as a specific probe.

Furthermore, the gene or cDNA may be used for the transfection of cells, which transfection results in cells expressing or secreting the desired glypcan. The transfected cells can be used to produce transgenic animals therefrom, which in case the gene is an aberrant gene, can be used to study the effect of the aberration or to test medicaments. Alternatively, natural glycans may be isolated or recombinant (wild type or mutated) glycans produced (in transfected cells or transgenic animals) for use as therapeutics. Such therapeutics may be used to mimic the biological effects of the glycans (control of cell growth and differentiation),

in attempts to remedy the effects of absolute or relative deficiencies of these genes or to enhance the effects of the normal genes. With the appropriate modifications of the glypcan gene sequences, therapeuticals based on 5 modified glypcan gene sequences may also be used to block the effects of the glypcans, in attempts to remedy the effects of absolute or relative overexpressions or activities of the products of the various glypcan genes.

As a non-limiting example, recombinant soluble 10 glypcans may be used as decoy receptors for antagonizing the effects of factors that depend on membrane-anchored glypcans, whereas the delivery of membrane-intercalatable glypcans to cells may restore cellular sensitivity to these factors.

15 Diagnostic methods according to the invention comprise but are not limited to the following.

A method for diagnosing aberrations in a glypcan encoding gene, comprises isolation of the gene from cells expected to be harboring an aberrant gene; and 20 comparison of the nucleotide sequence of the gene thus obtained with the nucleotide sequence of a wild type gene. The term "wild type gene" as used in this application is intended to encompass a gene from a non-affected individual. The sequences given in figures 1 to 25 5 are representatives for wild type gene sequences.

Comparison between the potentially aberrant and wild type nucleotide sequences can be performed at various levels. On a first level it can be established whether the expected aberration(s) has (have) resulted in 30 restriction fragment length polymorphism. In order to do this the isolated gene and a wild type comparison gene are separately digested with one or more selected restriction enzymes. The digest thus obtained is separated on a gel revealing a pattern of bands. 35 Differences in the pattern indicate the presence of differences in the restriction sites present in the polynucleotide and thus changes in the sequence thereof.

Deletions can be detected by means of any nucleic acid amplification technique such as the Polymerase Chain Reaction (PCR). For this, probes are identified corresponding to various parts of the gene to 5 be diagnosed, for example exons. Amplification between a set of probes will only occur if the part of the gene to which a selected set of probes should hybridize is still present. In addition or as an alternative the length of the amplified fragment indicates whether any part is 10 deleted.

Point mutations can be identified by more sophisticated techniques such as SSCP (single-strand conformation polymorphism screening), heteroduplex analyses, DNA-chips, chemical and enzymatic methods, 15 sequencing of PCR products, denaturant gradient gel electrophoresis or other state of the art methods that may become available in the future.

Other diagnostic methods comprise the in situ detection of physical changes like translocations, 20 inversions or deletions. Translocations can be detected by hybridizing a set of chromosomes with a first probe that hybridizes to a part of the glypcan gene that is not likely to be involved in the translocation, inversion or deletion and a second probe that hybridizes to a part 25 of the glypcan gene that is likely to be involved in such aberration. When a translocation has occurred the second probe will be found on another chromosome than the first one. If the probable translocation partner is identified, an additional set of probes can be used which 30 hybridize to the translocated part and the remaining part, respectively of the translocation partner and bearing a different label from the first set of probes. Upon translocation one of the probes of the first set will be found on the chromosome of the translocation 35 partner and one probe of the second set will be found on the chromosome of the glypcan gene and vice versa.

Identifying inversions and deletions works in a similar way with two probes, one that hybridizes to a

part of the gene that is not likely to be involved in the inversion or deletion and a second probe that is likely to be involved in such aberration. In case of an inversion the second probe will be found closer to or 5 further away from the first probe than in a non-aberrant chromosome. In the case of deletions one of the probes will be missing on the aberrant chromosome. The description and examples of this application give various examples of "parts of the gene that are likely to be 10 involved in an aberration".

Diagnosis can also be performed at the level of the (potentially absent or aberrant) protein encoded by the glypcan gene. Antibodies directed to the gene product or protein can be used on Western blots to detect 15 the presence of the protein in the cell or to assess the amount of protein present.

The diagnostic tests of the invention can be performed on various source materials. RFLP, deletion PCR, SSCP and chromosome analyses are for example 20 performed on blood cells or tissue biopsy samples of the patient and his or her family. Furthermore, tumor cells and normal cells of these subjects may be used. For protein analysis, tissue samples, sera, tissue fluids of patients and family, pleura exudates, ascites etc. may be 25 used.

All these diagnostic methods are based on the information that can be derived from the various genes and the gene products thereof. This information is given in the following figures.

30 Figure 1 shows the nucleotide sequence of the glypcan-6 cDNA, comprising the coding sequence of the newly identified GPC6 gene, and the predicted amino acid sequence.

Figure 2 shows the nucleotide sequence of the 35 glypcan-4 cDNA, comprising the coding sequence of the GPC4 gene, and the predicted amino acid sequence.

Figure 3 shows the nucleotide sequence of the human glypcan-1 cDNA, comprising the coding sequence of

the GPC1 gene, and the predicted amino acid sequence (Genbank accession number X54232; David et al., J. Cell Biol. 111, 3165-3176, 1990).

Figure 4 shows the nucleotide sequence of the 5 human glypcan-3 cDNA, comprising the coding sequence of the GPC3 gene, and the predicted amino acid sequence (Genbank accession number Z37987).

Figure 5 shows the nucleotide sequence of the human glypcan-5 cDNA, comprising the coding sequence of 10 the GPC5 gene, and the predicted amino acid sequence (Genbank accession number U66033; Veugelers et al. Genomics 40, 24-30, 1997).

Figure 6 shows the alignment of the predicted amino acid sequences of the members of the glypcan 15 family. GPC1 is human glypcan (David et al., 1990), GPC3 is the translated ORF of MXR7 (GenBank #Z37987) and the human homologue of rat OCI-5, GPC4 is human glypcan-4 (see example 2) and the human homologue of K-glypcan (Watanabe et al., 1995), GPC5 is human glypcan-5 20 (Veugelers et al., 1997). The set of fourteen conserved cysteines and the putative glycosaminoglycan attachment sites are outlined by underlining. Serines occurring in SGXG sequence contexts are indicated in bold. Alignment was done using the Clustal V program (Higgins, 1994). 25 Single carets under the sequences indicate exon-intron boundaries occurring within codons; double carets indicate exon-intron boundaries occurring between codons.

In the following examples reference is made to the following additional figures.

30 Figure 7A shows a Northern blot for GPC6 of human fetal Brain (lane 1), Lung (lane 2), liver (lane 3) and Kidney (lane 4) RNA. The positions of RNA size markers (kb) are indicated in the abscissa.

Figure 7B shows a Northern blot for GPC6 of 35 human adult Heart (lane 1), Brain (lane 2), Placenta (lane 3), Lung (lane 4), Liver (lane 5), Skeletal Muscle (lane 6), Kidney (lane 7) and Pancreas (lane 8) RNA.

Figure 7C shows a Northern blot for GPC6 of human adult Spleen (lane 1), Thymus (lane 2), Prostate (lane 3), Testis (lane 4), Ovary (lane 5), Small Intestine (lane 6), Colon Mucosal lining (lane 7), and 5 Peripheral Blood Leukocyte (lane 8) RNA. The positions of RNA size markers (kb) are indicated in the abscissa.

Figure 8 shows heparan sulfate core protein expression in control and GPC-transflectant Namalwa cells. Western blotting of non digested, heparitinase-digested, 10 doubly heparitinase- and chondroitinase ABC-digested, and chondroitinase ABC-digested proteoglycan fractions, using the delta-HS-specific mAb 3G10. This antibody reacts with the desaturated uronates that are generated by heparitinase and that remain in association with the core 15 protein after the enzyme treatment and during electrophoresis. After heparitinase it therefore detects all heparan sulfate proteoglycan core proteins present in the sample. The positions of protein size markers are indicated in the abscissa. Hase: heparitinase; Case: 20 chondroitinase ABC. Control: wild type Namalwa cells (wt) and Namalwa cells, transfected with pREP4 Ø; GPC1, GPC4 and GPC6: Namalwa cells, transfected with respectively the plasmids glyp1-pREP4, glyp4-pREP4 and glyp6-pREP4.

Figure 9 shows heparan sulfate expression in 25 control and GPC-transflectant Namalwa cells. FACS analysis of non-digested and heparitinase-digested cells, using the native HS-specific 10E4 antibody (non digested cells (----)) and delta-HS-specific 3G10 antibody (digested cells (-----)). Control: Namalwa cells, transfected with 30 pREP4; GPC1, GPC3, GPC4, GPC5 and GPC6: Namalwa cells, transfected with respectively the plasmids glyp1-pREP4, glyp3-pREP4, glyp4-pREP4, glyp5-pREP4 and glyp6-pREP4.

Figure 10 shows the chromosomal localization of GPC6 to chromosome band 13Q32 as a photo (A) and a 35 schematic representation thereof (B). Arrows indicate colored bands.

Figure 11A shows the Northern blot for GPC4 of human fetal Kidney (lane 1), Liver (lane 2), Lung (lane

3) and Brain (lane 4) RNA. The upper part of the figures represents the hybridization with the GPC4 probe; the lower part the hybridization with a  $\beta$ -actin control probe.

5           Figure 11B shows the Northern blot for GPC4 of human adult Heart (lane 1), Brain (lane 2), Placenta (lane 3), Lung (lane 4), Liver (lane 5), Skeletal Muscle (lane 6), Kidney (lane 7) and Pancreas (lane 8) RNA.

10          Figure 11C shows the Northern blot for GPC4 of human adult Spleen (lane 1), Thymus (lane 2), Prostate (lane 3), Testis (lane 4), Ovary (lane 5), Small intestine (lane 6), Colon; mucosal lining (lane 7), peripheral blood leukocyte (lane 8). The positions of RNA size markers (kb) are indicated in the abscissa.

15          Figure 12A illustrates the chromosomal localization of GPC4 to chromosome Xq26 and relative order of GPC3 and GPC4. For initial chromosomal localization of GPC4, FISH was performed using either BAC 35H9 or BAC 68G14 on metaphase spreads, prepared from 20 PHA-stimulated normal peripheral blood leukocytes (Figure 12A). For relative ordering of GPC genes FISH was performed with BAC's for GPC4 (35H9, 68G14 labeled in red) and BAC's for GPC3 (166D10 and 36D20, labeled in green) on PHA-stimulated cell lines GM3884, GM13034 and 25 GM0097 (Figures 12B, 12C and 12D). Chromosomes were counterstained with DAPI, and the images were taken using a cooled CCD device. Arrows indicate the positive signals at chromosome Xq26.

30          Figure 13A and 13B show a BAC/PAC contig linking GPC4 to GPC3 on Xq26. Figure 13C shows glycan deletions found in SGBS-patients. STS's are indicated by black circles; exons are indicated by grey squares. Not drawn to scale (The distance between SWXDI698 and exon-8 of GPC3 is approximately 250 kb (See Pilia et al., 1996).

35          The following tables that precede the references give inter alia suitable primers for use in the various diagnostic methods:

Table 1 shows the primers used in 5'-RACE experiments for the identification of GPC6.

Table 2 shows the percentages of amino acid identities between glypicans.

5 Table 3 shows the primers used in the RACE experiments with GPC4.

Table 4 shows the gene specific primers used for sequencing of the GPC4 gene.

Table 5 shows the novel STSs MV1, MV2 and MV3.  
10 Table 6 shows localization of FISH signals in SGBS patients.

Table 7 shows the intron-exon organization of GPC4.

Table 8 shows primers to be used in deletion analysis of GPC3 and GPC4.

15 Table 9 shows the primers for use in SSCA of GPC4.

Table 10 shows the results of deletion and SSCP screening in 8 patients with SGBS.

20 Table 11 shows primer pairs for deletion PCR of GPC5.

Table 12 shows primer pairs for deletion PCR of GPC6.

The present invention will be further  
25 illustrated in the accompanying examples that are in no way intended as a limitation of the present invention.

#### EXAMPLES

##### 30 EXAMPLE 1

###### Identification and characterization of glycan-6

###### 1.1 Introduction

The isolation of one of the cDNAs of the present invention (GPC6) started from EST database  
35 entries showing significant similarity with (cDNA coding for) glypicans. The cDNA was found in a cDNA library of fetal brain.

1.2 Materials and methods

## 1.2.1 Bioinformatics

EST entries (including homology data) were retrieved from dbEST using a text string based query 5 interface (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>). Protein alignments were made using the program Clustal. DNA alignments were made using the program GENEPRO.

1.2.2 Molecular cloning of human GPC6

10 The primer sets used for the 5'-RACEs (5'rapid amplification of cDNA ends), corresponding to two GPC-like ESTs (GenBank Accession Nos. N87558 and AA001322), are given in Table 1. The cDNAs were amplified from a library of adaptor-ligated double strand human fetal 15 brain cDNA (Clontech, Palo Alto, CA) through a two-step PCR protocol. In the first PCR a gene-specific primer was used and an anchor primer provided by the supplier. Then 1 µl of each first PCR was used as template for a second PCR, using a second gene-specific nested primer (cf. 20 Table 1) and a nested anchor primer provided by the supplier. The products of the second steps were analyzed by electrophoresis in a 0.6% agarose gel. Distinct bands were gel purified using the QIQUICK II DNA clean-up System (Qiagen), T/A-cloned in the plasmid pCR2.1 25 (Invitrogen), and sequenced using a Pharmacia A.L.F. DNA Sequencer, with Dye Primer Cycle Sequencing chemistry on double stranded plasmid templates. In total, 5 independent clones from two separate 5'-RACE experiments (2 from the 5'-RACE-1 and 3 from the 5'-RACE-2 30 experiment) were sequenced.

Clone zh83a06 from the Soares fetal liver/spleen library (Lennon et al., 1996), which had yielded EST No. AA001322, was obtained from the I.M.A.G.E. Consortium 35 (<http://www-bio.llnl.gov/bbrp/image/image.html>) through Research Genetics, Inc. (Huntsville, AL). This clone (ID: 427858) was completely sequenced, yielding residues 1835-

2748 of the composite cDNA sequence that is shown in figure 1.

1.2.3 BAC cloning and chromosomal localization of GPC6 by  
5 FISH

To isolate genomic clones for GPC6, part of the cDNA insert from I.M.A.G.E.-clone 427858 was isolated with EcoRI/PstI, gel purified, labeled and used to screen a human genomic BAC library (Research Genetics, Inc., 10 Huntsville, AL). Two BACs, 114A17 and 182F5, were isolated. Their authenticity was verified by Southern blotting. These results indicated that the BACs 114A17 and 182F5 contain exons 6-9 of GPC6. BAC DNA was labeled with bio-16-dUTP (Sigma) by nick-translation using a 15 commercial kit (Life Technologies, Gaithersburg, MD).

Metaphase spreads were prepared from PHA-stimulated human peripheral blood lymphocytes cultured for 72 hours. Prior to FISH, slides were treated with RNAse A and pepsin as described (Wiegant et al., 1991). 20 Human Cot1 DNA (Life Technologies) was used as a competitor. Denaturation of the slides and probes, hybridization, and subsequent cytochemical detection of the hybridization signals were performed as previously described (Vermeesch et al., 1995). Chromosomes were 25 counterstained with DAPI and the slides were mounted in Vectashield mounting medium (Vector Laboratories Inc, Burlingame, CA). The signal was visualized by digital imaging microscopy using a cooled charge-coupled device camera (Photometrics Ltd, Tucson, AZ). Merging and 30 pseudocoloring were performed using the Smart Capture software (Vysis, Stuttgart, Germany).

1.2.4 Northern blotting

The membranes for the Northern blots were 35 obtained from Clontech. Hybridization was performed for two hours at 68°C, using Expresshyb solution (Clontech) according to the manufacturer's specifications. The probe was either a <sup>32</sup>P-oligolabeled BamHI-XbaI fragment from the

I.M.A.G.E.-clone 427858 (corresponding to residues 2147-2488 of the GPC6 sequence) or a HindIII-BamHI fragment from the GPC6 composite cDNA sequence (corresponding to residues 1724-2147 of the GPC6 sequence). Dehybridisation 5 included two washes with 2.0% SSC, 0.05% SDS (5 min at RT; 30 min at 60°C) and a high stringency wash with 0.1% SSC, 0.1% SDS (30 min at 65°C).

1.2.5 Construction of expression plasmids and cell  
10 transfection

The NotI-BstEII and BstEII-AvaI fragments from overlapping RACE-clones, and the AvaI-HindIII fragment of I.M.A.G.E.-clone 427858 were ligated together in pCR2.1. NotI-XbaI and XhoI-XbaI fragments from this construct, 15 containing the Kozak sequence and initiator ATG, the full coding sequence and the stop codon, were subcloned in, respectively, pcDNAlII and pBluescript. For episomal expression in Namalwa cells, the full length cDNA was released from pBluescript with KpnI and NotI, and ligated 20 into pREP4, yielding the plasmid glyp6-pREP4.

Namalwa cells (ATCC CRL 1432) were routinely grown in DMEF12 medium supplemented with 10% FCS and L-glutamine. For transfection, the cells were prewashed with Ca- and Mg-free PBS and incubated for 10 min at 4°C 25 (10<sup>7</sup> cells in 1 ml Ca/Mg free PBS) with 30 µg glyp6-pREP4 plasmid before electroporation at 240 V and 960 µF (Gene Pulser, Biorad).

Selection was started 48 h later with 250 µg/ml of hygromycin B. Stable transfection was achieved after 30 12 days. Expression of heparan sulfate in the transfectants was analyzed by FACS, using the HS-specific antibody 10E4 and the delta-HS-specific antibody 3G10, and by Western blotting, using the 3G10 antibody as described before (David et al., 1992; Steinfeld et al., 35 1996). Stable expressing transfectant Namalwa clones were isolated from cells transfected with linearised glyp6-pcDNAlII, selected in media containing 400 µg/ml of G418, and panned on 10E4 antibody. (The HS-specific antibodies

10E4 and 3G10 have been isolated, characterized and produced in our own laboratory (David et al., 1992). They are now also commercially available from Seikagaku Co.)

5 1.3 Results

1.3.1 Molecular cloning of glypican-6 (GPC6), a novel glypican

During the screening of public EST databases, it was noticed that there was one EST from human fetal heart (GenBank Acc. No. N87558; clone not available) and one EST from fetal liver/spleen (GenBank Acc. No. AA001322; available as clone 427858 from the I.M.A.G.E.-consortium) with very high homology (but not identical) to human GPC1 and GPC4 (approximately 70% identity at the nucleotide- and encoded amino acid-sequence level). It was assumed that these might represent (a) novel glypican(s), and primers (annealing to regions with significant sequence divergence from GPC1 and GPC4) were designed which would amplify the corresponding cDNA(s) (see Table 1). These primers were used on the same human fetal brain cDNA library that was used for the isolation of GPC5 (Veugelers et al., 1997) and human GPC4 (example 2), in RACE experiments.

From the analysis of their sequences, it appeared that all clones from these RACE experiments and the I.M.A.G.E.-clone 427858 represented overlapping cDNAs.

Figure 1 represents the merged sequences of these clones, and the predicted structure of the protein encoded by the message that corresponds to this cDNA. The sequence features an ATG start codon, in a Kozak sequence context, at position 586 and a TAA stop codon at position 2251. Two AATAAA sequences (potential polyadenylation signals) are present at positions 2598 and 2690. The open reading frame in the sequence codes for a protein of 555 residues. The protein sequence starts and terminates with hydrophobic signal peptide-like sequences. It contains no asparagines that correspond to potential N-glycosylation

sites, and contains four serine-glycine dipeptide sequences. All four Ser-Gly dipeptide sequences occur towards the C-terminus of the protein, and form part of a direct Ser-Gly repeat sequence. This Ser-Gly tetra-repeat 5 sequence is flanked, both upstream and downstream, by acidic amino acids (D/E), reproducing a motif that has been reported to promote the assembly of heparan sulfate in proteoglycans. The downstream acidic residues occur within the sequence CMDDVC, and may reproduce a motif (a 10 small acidic loop supported by a disulfide bond) that is shared by most glypicans (except glypican-2). This loop follows the SG repeats in the glypicans 1, 4 and 6, but interrupts or precedes the SG repeats in the glypicans 3 and 5.

15         Alignment of this predicted protein sequence with the protein sequences of the other known members of the glypican family (glypicans 1, 3, 4, and 5 as identified in man, and glypican-2 as identified in rat) revealed significant sequence similarities (figure 6).  
20 This similarity included the 14 cysteines and the position and identity of several additional amino acid residues that are conserved in all glypicans identified so far. The entire protein showed 63% of sequence identity to human glypican-4, 44% of identity to human 25 glypican-1, and 24-25% identity to the human glypicans 3 and 5. Comparison with rat glypican-2 showed only 41% of identity (Table 2). Since (where both available) human and rodent glypican sequences had always proven highly similar (~90% of sequence identity), it seemed unlikely 30 that this protein represented the human homologue of cerebroglycan (glypican-2). This protein was therefore designated as glypican-6, the sixth member of the vertebrate glypican family.

35         This alignment further indicated that the glypicans 4 and 6 were more closely related to one another (63% identity) than to the other glypicans (only 20 - 40% identity), and that the glypicans 3 and 5 were more closely related to one another (43% identity) than

to the other glypicans (~20% identity) (see Table 2). The high similarity of glypican-4 and glypican-6 became even more striking when the N-terminal and C-terminal hydrophilic signal peptide-like sequences (absent from the 5 mature proteins) were excluded in the alignments (identity raising to 70%). These data suggest that the glypican family of cell surface heparan sulfate proteoglycans (as known today) may be composed of discrete subfamilies: one comprising glypicans 4 and 6, 10 and possibly also glypican-1 (and 2); the second comprising glypicans 3 and 5.

### 1.3.2 Expression of glypican-6

In Northern blotting experiments (figures 7A, 15 7B and 7C), two different GPC6 probes (one from the 3' UTR, corresponding to residues 2147-2488 of the sequence shown in figure 1, the other corresponding to residues 1724-2147 of the sequence shown in figure 1) detected a transcript of ~7kb in (all) fetal tissues (analyzed 20 here). High levels of expression were apparent in fetal kidney, moderate levels in fetal lung and fetal liver, and a low level of expression in fetal brain (figure 7A). In adult tissues the message appears to be expressed almost ubiquitously. The message is expressed at very 25 high levels in ovary, and at high levels in liver, kidney, small intestine and colon (mucosal lining). The message is also present at low levels in heart, brain, placenta, lung, skeletal muscle, pancreas, spleen, thymus, prostate and testis (figures 7B and C). The 30 message is undetectable in peripheral blood leukocytes. In adult kidney the probes also detected a second less abundant message of approximately 5.8 kb. Adult heart and adult skeletal muscle yielded an extra band of ~3.9 kb.

These data indicated differential and only 35 partially overlapping expressions of the GPC6 and other GPC messages in the different tissues, further evidence that the various GPCs are distinctive transcripts.

1.3.3 Identification of glypican-6 as a heparan sulfate  
proteoglycan

To test if glypican-6 would drive the synthesis of heparan sulfate, the glypican-6 insert was subcloned 5 in the pREP4 episomal expression vector and transfected in Namalwa cells. The Namalwa cells used for these experiments had previously been shown to express little endogenous heparan sulfate, but to support the synthesis of large amounts of heparan sulfate when transfected with 10 cDNAs (cloned in pREP4) that code for syndecans or glypican-1. Analysis of heparitinase digested proteoglycan from transfectant and control cells by Western blotting, using the delta-HS-specific mAb 3G10, confirmed that the transfectant cells expressed heparan 15 sulfate proteoglycan core proteins of ~65 kDa (major band) and ~18-14 kDa (minor bands, possibly metabolic degradation products) that were not detectable in the control cells (figure 8). Major bands of ~65 kDa and minor bands of smaller sizes were also observed for 20 transfectant Namalwa cells expressing glypican-1 or glypican-4 (figure 8) and glypican-3 or glypican-5 (not shown). FACS analyses of these cells with the HS-specific mAb 10E4 and, after heparitinase, with mAb 3G10 revealed a dramatic increase in the expression of cell surface-HS 25 in the transfectants (figure 9).

1.3.4 Chromosomal mapping of GPC6

Two BACs for GPC6, 114A17 and 182F5 were used to localize GPC6 to chromosome band 13q32 by fluorescent 30 in situ hybridization on metaphase chromosomes (figure 10; BAC 114A17). From this it follows that GPC5 (closely related to GPC3) and GPC6 (closely related to GPC4) map in close proximity of one another on 13q32, mimicking the clustering of the GPC3 and GPC4 genes on chromosome Xq26 35 (see example 2).

**EXAMPLE 2**Identification and characterization of human glypican-42.1 Introduction

The isolation of the cDNA that is used in the  
5 present invention for diagnostics (GPC4) started from a  
partial cDNA for human glypican-4. A cDNA comprising the  
complete coding sequences for human glypican-4 was found  
in a cDNA library of fetal brain.

10 2.2 Materials and methods2.2.1 Bioinformatics

EST entries (including homology data) were  
retrieved from dbEST using either a text string based  
query interface

15 (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>), or by  
BLAST searches using the BLAST-server  
(<http://www.ncbi.nlm.nih.gov/BLAST/>). Protein alignments  
were made using the program Clustal (Higgins et al.,  
1994). DNA alignments were made using the program  
20 GENEPRO.

2.2.2 Molecular cloning of human GPC4

A partial cDNA for human GPC4 was obtained by  
PCR on a human fetal kidney library (pKGP-PCR). The  
25 sequence of this cDNA was used to design the primers for  
the RACE experiments and the isolation of cDNA for the  
complete coding sequence of human GPC4. The 5'-RACE and  
3'-RACE experiments were performed on a library of  
adaptor-ligated ds fetal brain cDNA, using the Marathon  
30 cDNA Amplification kit from Clontech (Palo Alto, CA). The  
cDNAs were amplified through a two-step PCR protocol. The  
first PCR used a gene-specific primer (Table 3) and an  
anchor primer provided by the supplier. Then 1 µl of the  
first PCR reaction was used as template for the second  
35 PCR-reaction, using a second gene-specific nested primer  
and a nested anchor primer provided by the supplier. The  
products of the second PCR were analyzed by  
electrophoresis in a 0.6% agarose gel. Distinct bands

were gel purified using the Qiaquick DNA clean-up system (Qiagen), T/A-cloned in the plasmid pCR2.1 (Invitrogen), and sequenced using a Pharmacia A.L.F. DNA Sequencer, with Dye Primer Cycle Sequencing chemistry on double 5 stranded plasmid templates. In total 3 independent RACE clones were sequenced for the 5'-RACE and 3 independent RACE clones were sequenced for the 3'-RACE.

Additionally, clone zx12d12 from the Soares' 9 week normal fetus cDNA library (Lennon et al., 1996) was 10 obtained from the I.M.A.G.E. Consortium (<http://www.bio.llnl.gov/bbrp/image/image.html>) through Research Genetics, Inc. (Huntsville, AL). This clone (ID: 786263) was completely sequenced, yielding residues 1443-2315 of the composite cDNA sequence shown in figure 2. 15 All the sequences obtained for the coding region (residues 213-1883) were derived from at least two different RACE-products.

#### 2.2.3 BAC cloning and chromosomal localization of GPC4 by 20 FISH

To isolate genomic clones for GPC4, the pKGP-PCR probe (corresponding to residues 422-1497 of the GPC4 cDNA sequence shown in figure 2) and the NotI-BglII fragment (residues 1-386) of the GPC4 cDNA were gel 25 purified, <sup>32</sup>P-labelled and used to screen a human genomic BAC library (Research Genetics, Inc., Huntsville, AL). Two BACs, 35H9 and 151D8 were isolated with the PCR probe, and one BAC, 68G14, with the NotI-BglII fragment. The authenticity of these clones was verified by Southern 30 blotting and by cycle-sequencing of the exon-intron boundaries, using gene-specific primers derived from the GPC4 cDNA sequence (Table 4). BAC DNA was labeled with bio-16-dUTP (Sigma) by nick-translation, using a commercial kit (Life Technologies, Gaithersburg, MD).

35 A similar strategy was used to isolate BACs for GPC3. Using cDNAs corresponding to residues 1-2300 and 1-408 of the GPC3 sequence (Genbank Access N° Z37987), two

BACs were identified: BAC 166D10, which contained exon-3 of GPC3, and BAC 36D20 which contained exon-2 of GPC3.

Metaphase spreads were prepared from PHA-stimulated human peripheral blood lymphocytes cultured 5 for 72 hours. Prior to FISH, slides were treated with RNase A and pepsin as described (Wiegant et al., 1991). Human Cot1 DNA (Life Technologies) was used as a competitor. Denaturation of the slides and probes, hybridization, and subsequent cytochemical detection of 10 the hybridization signals were performed as previously described (Vermeesch et al., 1995). Chromosomes were counterstained with DAPI and the slides were mounted in Vectashield mounting medium (Vector Laboratories Inc, Burlingame, CA). The signal was visualized by digital 15 imaging microscopy using a cooled charge-coupled device camera (Photometrics Ltd, Tucson, AZ). Merging and pseudocoloring were performed using the Smart Capture software (Vysis, Stuttgart, Germany).

20 2.2.4 GPC4 gene structure and BAC/PAC contig of the GPC3/GPC4 gene cluster on Xq26

Exon-intron boundaries were determined by cycle-sequencing of BAC DNA using gene specific primers. Alternatively, BAC DNA was subcloned in plasmids, 25 verified for the presence of GPC4 exons (by PCR and Southern blotting) and subsequently sequenced.

YAC's yWXD363, yWXD2789-I, yWXD440, yWXD736, yWXD69, yWXD808, yWXD6857-I, yWXD6858-I, yWXD3373, yWXD2704-I, yWXD6142, yWXD2724-I from the Xq26 contig 30 (Pilia et al., 1996) were obtained from the American Type Culture Collection (ATCC) and verified for GPC4 content by PCR and Southern blotting.

The ends of BAC's 35H9 and 68G14 were sequenced and used to construct the novel sequence-tags (STS) MV1, 35 MV2 and MV3 (Table 5).

The PAC-library from P. de Jong was screened with <sup>32</sup>P-oligonucleotide probes for MV1, exon-1 GPC4, exon-2 GPC4 and exon-8 GPC3. PAC content was verified by PCR.

STS's for GPC4 exons are described in Table 5, STS's sWXD1698, sWXD1165 and sWXD2342 have been described by others (see The Genome Database <http://www.gdb.org>). The reaction cycles for STS's MV1, MV2 and MV3 were: 94°C for 5 30 sec, 55°C for 30 sec, 72° for 30 sec, for 35 cycles. Cycling was preceded by a 150 sec incubation at 94°C.

#### 2.2.5 Northern blotting

The membranes for the Northern blots were obtained from Clontech. Hybridization was performed for two hours at 68°C, using Expresshyb solution (Clontech) according to the manufacturer's specifications. The probe was either a <sup>32</sup>P-oligolabeled NotI-BglII fragment from one of the 5'RACE clones (corresponding to residues 1-386 of the sequence shown in figure 2), or a <sup>32</sup>P-oligolabeled BamHI-BamHI fragment from the composite cDNA sequence constructed in pREP4 (corresponding to residues 1148-2291 of the sequence shown in figure 2). Dehybridisation included washing at room temperature for 30 min with 2.0% SSC, 0.05% SDS and a high stringency wash for 30 min at 0.1% SSC, 0.1% SDS and 65°C.

#### 2.2.6 Mutational analysis of the GPC4 gene

From the characterization of the corresponding intron/exon boundaries in GPC4, primer pairs were designed for the amplification of all exons of the human GPC4 gene, to permit deletion and mutational analysis (Table 8). For GPC3 deletion analysis, we designed new primers for the amplification of all exons (Table 8). Genomic DNA was obtained from one newly identified patient, counseled at the Center for Human Genetics (CME) of the University of Leuven (with informed consent from the parents); from the lymphoblastoid cell lines AG0817, AG0857, AG0893, AG0946, AG0969, and FY0367 (database IDs 30 from the European Collection of Cell Cultures (ECACC); and from the fibroblastic cell lines GM13034, GM3884, GM0097 (ATCC), all established from patients with SGBS. All patient DNAs were analyzed by PCR. The reaction

cycles were: 94°C for 30 sec, annealing temperature for 30 sec, 72°C for 30 sec, for 35 cycles. Cycling was preceded by a 2.5 minute incubation at 94°. The reaction products were analyzed by electrophoresis in 2% agarose gels or, alternatively, were analyzed for single-strand conformation polymorphisms (SSCP) in non-denaturing polyacrylamide gels as described previously (Matthijs et al., 1997). PCR products with variant SSCs and controls were sequenced, either directly after gel purification or 10 after T/A cloning in pCR2.1 (Invitrogen). In the latter case, several independent clones from independent amplifications were characterized by Dye Primer Cycle Sequencing.

15 2.2.7 Construction of expression plasmids and cell transfection

To construct a partial GPC4 cDNA, containing the entire coding region, a NotI-EcoRI fragment from a 5' RACE clone, a EcoRI-PstI fragment from pKGP, and a PstI-BamHI fragment from a 3'-RACE clone were ligated together in pBluescript. A NotI-NotI fragment containing GPC4 was isolated from this construct and ligated in pCDNAIII and pREP4, yielding respectively glyp4-pcDNAIII and glyp4-pREP4. Namalwa cells (ATCC CRL 1432) were routinely grown 25 in DMEM12 medium supplemented with 10% FCS and L-glutamine. For transfection, the cells were prewashed with Ca- and Mg-free PBS and incubated for 10 min at 4°C ( $10^7$  cells in 1 ml Ca/Mg free PBS) with 30 µg glyp4-pREP4 plasmid before electroporation at 240 V and 960 µF (Gene 30 Pulser, Biorad). Selection was started 48 h later with 250 µg/ml of hygromycin B. Stable transfection was achieved after 12 days. Expression of heparan sulfate in the transfected cells was analyzed by FACS, using the HS-specific antibody 10E4 and the delta-HS-specific antibody 35 3G10, and by Western blotting, using the 3G10 antibody as described before (David et al., 1992; Steinfield et al., 1996). Stable expressing transfected Namalwa clones were isolated from cells transfected with linearized glyp4-

pcDNAIII, selected in media containing 400 µg/ml of G418, and panned on 10E4 antibody.

### 2.3 Results

#### 5 2.3.1 Molecular cloning of human glypican-4 (GPC4)

The combination of 5'-RACE and 3'-RACE experiments, performed on a library of adaptor-ligated ds fetal brain cDNA library, yielded the complete coding sequence for human GPC4. Figure 2 represents the merged 10 sequences of the RACE clones, pKGP and EST zx12d12 (identified as GPC4 from BLAST searches) of public databases and the predicted structure of the protein encoded by the message that corresponds to this cDNA. The sequence features an ATG start codon, preceded by a Kozak 15 sequence, at position 213 and a TAA stop codon at position 1881. One AATAAA sequence (potential polyadenylation signal) is present at position 3697, and a stretch of polyA starts at position 3706. The predicted amino acid sequence of human GPC4 (556 residues) was 20 found to be highly homologous to that of mouse GPC4 ( $\text{K-glypican}$ ) (93.5% sequence identity) (see also Table 2).

The protein sequence starts and terminates with hydrophobic signal peptide-like sequences. It contains 25 three serine-glycine dipeptide sequences. All three Ser-Gly dipeptide sequences occur towards the C-terminus of the protein, and two of these form part of a direct Ser-Gly repeat sequence. These Ser-Gly sequences are flanked, both upstream and downstream, by acidic amino acids 30 (D/E), reproducing a motif that has been reported to promote the assembly of heparan sulfate in proteoglycans (Zhang et al., 1995). Because of the presence of three Ser-Gly repeats, glypican-4 would be predicted to have up to three heparan sulfate chains implanted on its core 35 protein. The acidic residue downstream of the Ser-Gly repeat occurs within the sequence CEYQQC, and may reproduce a motif (a small acidic loop supported by a disulfide bond) that is shared by most glypicans (except

glypican-2). This loop follows the SG repeats in the glypicans -1, -4 and -6, but interrupts or precedes the SG repeats in the glypicans -3 and -5.

5 2.3.2 Expression of human glypican-4

In Northern blotting experiments, both probes corresponding either to residues 1-386, or residues 1148-2291 of the GPC4 sequence were detecting two messages, one of 2.9 and one of 4.3 kb. The messages were 10 expressed, in several, but not all, of the human fetal and adult tissues tested (see figure 11). The origin of these two bands is not known, but could be due to the alternative usage of multiple polyadenylation signals, alternative splicing or, less likely, cross-hybridization 15 with messages for other (possibly yet to be identified) members of the glypican gene family. In fetal tissues the messages were expressed in brain, kidney and lung; but barely detectable in liver. In adult tissues the message is highly abundant in skeletal muscle, pancreas, kidney, 20 placenta, lung, heart, spleen, testis, ovary, colon, small intestine. Less intense bands were seen in brain, thymus and prostate, and barely detectable bands were seen in the liver. The message appears to be absent from peripheral blood leukocytes. EST's for human GPC4 were 25 also present in libraries prepared from a 9 week old fetus, pregnant uterus, fetal heart, adult lung, placenta and colon. The expression pattern of human GPC4 is almost the same as murine K-glypican with the exception of mGPC4 being abundantly expressed in the liver.

30

2.3.3 Identification of glypican-4 as a heparan sulfate proteoglycan

To test if glypican-4 would support the synthesis of heparan sulfate, the glypican-4 insert was 35 subcloned in the pREP4 episomal expression vector and transfected in Namalwa cells. The Namalwa cells used for these experiments had previously been shown to express little endogenous heparan sulfate, but to support the

synthesis of large amounts of heparan sulfate when transfected with cDNAs (cloned in pREP4) that code for syndecans or glypican-1. Analysis of heparitinase-digested proteoglycan from transfectant and controls 5 cells by Western Blotting, using the delta-HS-specific mAb 3G10, confirmed that the transfectant cells expressed heparan sulfate proteoglycan core proteins of ~65 kDa (major band) and ~18-14 kDa (minor bands) that were not detectable in the control cells (figure 8). FACS analyses 10 of these cells with the HS-specific mAb 10E4 and, after heparitinase, with mAb 3G10 revealed a dramatic increase in the expression of cell surface-HS in the transfectants figure 9).

#### 15 2.3.4 Chromosomal mapping of GPC4

Three BACs were identified for GPC4: BAC 35H9, BAC 151D8, and BAC 68G12. BACs 35H9 and 151D8 contained exons 2 to 9 of GPC4, while BAC 68G12 contained exon-1 of GPC4. FISH, performed on metaphase chromosomes, 20 localized all BACs for GPC4 to Xq26 (figure 12). Since GPC3 had also been localized to chromosome band Xq26 (Pilia et al., 1996), these results suggested that GPC3 (closely related to GPC5) and GPC4 (closely related to GPC6) probably mapped in proximity of one another on 25 Xq26, mimicking the clustering of the GPC5 and GPC6 genes on chromosome 13q32 (see example 1). The relative orientation of GPC4 and GPC3 was determined by FISH on cell lines of Simpson-Golabi-Behmel (SGBS) syndrome patients with translocations in the GPC3 gene (Table 6 30 and figure 12).

These FISH data indicate that the GPC4 gene lies centromeric to GPC3. Since there is a YAC-contig covering Xq26, it was decided to look for YAC's containing GPC4. The following YAC's were tested by 35 Southern-blotting and PCR for the presence of GPC4 exons: yWXD363, yWXD2789-I, yWXD440, yWXD736, yWXD69, yWXD808, yWXD6857-I, yWXD6858-I, yWXD3373, yWXD2704-I, yWXD6142-I, yWXD2724-I.

Only YAC's yWXD3373 and yWXD6858-I were found to be positive for exon- 2 to exon-9 of GPC4. No YAC's were found positive for exon-1 of GPC4. Moreover, only YAC's yWXD6142 and yWXD2704 were positive for exon-8 of 5 GPC3. These data suggested that some YAC's might have undergone internal deletions, and lead to the construction of a new BAC/PAC contig. Figure 13 shows the BAC/PAC contig containing the entire GPC4 gene and linking both GPC3 and GPC4. This contig indicates that 10 both glypcans form a tandem array with exon-1 and the promotor region of GPC4 lying adjacent to the last exon of GPC3. The GPC4 gene exon/intron structure is schematically shown in Table 7.

15 **EXAMPLE 3**

Glypican involvement in the Simpson-Golabi-Behmel syndrome

3.1 Introduction

Recently, deletions and translocations 20 involving the gene for glypican-3 (GPC3) have been shown to occur in patients with the Simpson-Golabi-Behmel overgrowth syndrome (Pilia et al., 1996). Not all patients with this X-linked condition, however, are affected by mutations of the GPC3 gene that can easily be 25 demonstrated (Lindsay et al., 1997). GPC4 was mapped by the present inventors on Xq26, in close proximity to GPC3, in an interval such that it would be deleted in at least one family with SGBS (Pilia et al, family c and figure 13C). Therefore, the possibility was investigated 30 of Xq mutations in patients with SGBS that affect GPC4, in addition to or rather than GPC3. The results show that in some patients this is indeed the case.

3.2 Materials and methods

From the characterization of the corresponding 35 intron/exon boundaries in GPC4 (analysis of the BACs described in example 2) and GPC3 new primers were designed for the amplification of all exons. Genomic DNA

was obtained from one newly identified patient, counseled at the Center for Human Genetics (CME) of the University of Leuven (with informed consent from the parents); from the lymphoblastoid cell lines AG0817, AG0857, AG0893, 5 AG0946, AG0969, and FY0367 (database IDs) from the European Collection of Cell Cultures (ECACC); and from the fibroblastic cell lines GM13034, GM3884, GM0097 (ATCC), all established from patients with SGBS. All patient DNAs were analyzed by PCR. The reaction cycles 10 were: 94°C for 30 sec, annealing temperature for 30 sec, 72°C for 30 sec, for 35 cycles. Cycling was preceded by a 2.5 minute incubation at 94°. Primers and annealing temperatures are given in Table 8. The reaction products were analyzed by electrophoresis in 2% agarose gels. PCR 15 primer pairs were designed for amplification of all exons of GPC4 (including exon/intron boundaries, Table 11) and the corresponding PCR products were analyzed for single-strand conformation polymorphisms (SSCP) in non-denaturing polyacrylamide gels as described previously 20 (Matthijs et al., 1997). PCR products with variant SSCs and controls were either directly sequenced after gel purification or T/A cloned in pCR2.1 (Invitrogen). Several independent clones from independent 25 amplifications were characterized by Dye Primer Cycle Sequencing.

### 3.3 Results

A summary of the PCR and SSCP analyses is given in Table 10. These analyses identified one patient with a 30 deletion that involved the entire GPC4 gene and part of the GPC3 gene (exons 7 and 8). No other GPC4 deletions were detected, but a partial deletion of GPC3 (exons 1 and 2) was also identified in the patient diagnosed at CME. SSCP of the GPC3 exons revealed polymorphism for 35 exon 3. Two of the patients with a variant exon 3 were brothers, and sequencing of the corresponding PCR products identified a T>A mutation leading to a missense mutation of R for W<sup>296</sup> in glycan-3 ((a) in Table 10).

Interestingly, W<sup>296</sup> corresponds to one of the residues that are strictly conserved in all glypicans identified so far. Deletion of one T nucleotide (del T875), leading to a frame shift mutation and termination, was the basis for 5 the variant SSC of exon 3 in a third patient (b). SSCA of the GPC4 exons revealed polymorphisms for the exons 7 and 8, in one and the same patient. Sequencing of the PCR product of exon-7 in this patient identified a G>T mutation leading to a substitution of D<sup>391</sup> by E in 10 glypican-4 (c). Sequencing of the PCR product of exon 8 in this patient and controls identified a C>T substitution leading to a substitution of V by A as residue<sup>442</sup> in glypican-4 (d). It may be noted that the RACE experiments also yielded V as residue 442 and D as 15 residue 391 (see figure 2), and that these residues have not been conserved in the glypicans. Moreover the IMAGE Consortium cDNA clone for GPC4 also had a V at position 442, and the plasmid pKGP had an E at position 391.

Table 1

Primers used in the 5'-RACE-experiments for GPC6

5'-ATTCCACTCTGTGTCGAGGTCAAGCCTGA-3' (1425-1452) : first primer (1)

5'-ACAGTATGGGCAGTACAGCATCTTCATG-3' (1329-1358) : nested primer (1)

5'-CCTGAGGCCACTGGATTCACTCACTTG-3' (2048-2072) : first primer (2)

5'-GCCATAATCTGCTGCTGATGAAAGTG-3' (1956-1982) : nested primer (2)

(The numbers in parenthesis refer to the corresponding residues in the composite cDNA sequence shown in figure 1).

Table 2

Percent amino acid identities between the glypicans

	GPC1	rGPC2	GPC3	GPC4	GPC5	GPC6
GPC1	100	35	21	42	21	44
rGPC2		100	19	37	17	41
GPC3			100	23	43	24
GPC4				100	24	63
GPC5					100	25
GPC6						100

Table 3

Primers used in the RACE-experiments for GPC4

5'-CGAGCCCAGAACGTCATTTAGCATTCTTCC-3': 5'-RACE GPC4  
 5'-CACAAACATATCATTCAGGGATTTCTCTGC-3': nested 5'-RACE GPC4  
 5'-ATTGAAGATCTGTCCCCAGGGTCTAC-3': 3'-RACE GPC4  
 5'-AGTGTGGTCAGCGAACAGTCAATC-3': nested 3'-RACE GPC4  
 5'-CCAACTGTGATCTGCCCTGTTCT-3': nested 3'-RACE GPC4

Table 4

PRO8-37	F-E	(AS)	GPC-4	Ex-1:	5'-ggggCATCgTTCTTgTTgAA
PRO7-44	F-E	(AS)	GPC-4	Ex-2:	5'-TCATCAAACCTCTTgTAACg
PR10-26	F-E	(AS)	GPC-4	Ex-3:	5'-AAGTGGTACTGGGAGTTAC
PR10-50	F-E	(AS)	GPC-4	Ex-4:	5'-CTCCAGCAAGCACAAATATG
PR06-40	F-E	(AS)	GPC-4	Ex-5:	5'-CTTCTgAgACACTTgAACAC
PR06-41	F-E	(AS)	GPC-4	Ex-6:	5'-CCAgTCggTCCAAACTAgTg
PR09-11	F-E	(AS)	GPC-4	Ex-8:	5'-AgTCCACgTCgTTCCCATTg
PR09-10	F-E	(AS)	GPC-4	Ex-9:	5'-CTCCACTCTCTCTgCATAAC
PR10-38	F-E	(AS)	GPC-4	Ex-9:	5'-CCAACTGTGATCTGCCCTGTTCT
PR10-12	F-E	(S)	GPC-4	Ex-1:	5'-TCCCgTCCggTCCCCAAAggT
PR10-13	F-E	(S)	GPC-4	Ex-2:	5'-CCTCTTgACTTCATgAAATg
PR10-14	F-E	(S)	GPC-4	Ex-3:	5'-TAAATgACTTCTgggCTCgC
PR06-38	F-E	(S)	GPC-4	Ex-3:	5'-TgCAgAgAAATCCCTgAAATg
PR10-44	F-E	(S)	GPC-4	Ex-4:	5'-ATGATCTACTGCTCCCTACTG
PR10-33	F-E	(S)	GPC-4	Ex-4:	5'-TGAGAAGAAATCACCTGCTC
PR10-15	F-E	(S)	GPC-4	Ex-5:	5'-ATTgTTCTCCCCAAACCTC
PR10-34	F-E	(S)	GPC-4	Ex-6:	5'-CTTCCCTAACATGACGACAC
PR06-39	F-E	(S)	GPC-4	Ex-7:	5'-ggTACTgATgTCAAggAgA
PR11-11	F-E	(S)	GPC-4	Ex-8:	5'-CGAATCATCTTACTGGGGTCA
PR11-38	F-E	(S)	GPC-4	Ex-9:	5'-CAGAACTGGGGTACAGCCTG
PR11-39	F-E	(S)	GPC-4	Ex-9:	5'-CTACATCTTAGTGGTGACCT

Table 5

5'-TCATGACTAGTTCTGCACGG-3': STS MV1a  
 5'-TGAAAATCCACATGATTGGAA-3': STS MV1b  
 5'-AAGCTTGAAGGGTGCCTCAGA-3': STS MV2a  
 5'-ATTCTCTGCTGCTGGTCACT-3': STS MV2b  
 5'-TCTCCTTCCCTGGACTAAC-3': STS MV3a  
 5'-TGAGTCAAATTAAAGAGCAAGGC-3': STS MV3b

Table 6

BAC (column) celline (row)	35H9 (GPC4)	68G14 (GPC4)	166D10 (GPC3)	36D20 (GPC3)
GM3884 t(X;16)	der (X)	der (X)	der (16)	der (16)
GM0097 t(X;1)	der (X)	der (X)	der (X)	der (X)
GM13034	Xq26	Xq26	Xq26	Xq26

Table 7

Exon	Intron 3' splice site	Exon 5' end	Exon sequence	Exon 3' end	Intron 5' splice site	Exon size	Intron number	Intron size (kb)
1	/	GCC	1-372	ACG	gtt	372	1	n.d.
2	cag	GTG	373-531	ATG	gtt	159	2	n.d.
3	cag	AAT	532-923	GTG	gtt	392	3	>4 kb
4	tat	GTA	924-1089	TAG	gtt	166	4	>2 kb
5	tag	ATG	1090-1220	AAG	gtt	131	5	0.105
6	cag	GTT	1221-1367	CTG	gtt	147	6	0.8
7	tag	GTT	1368-1504	CAG	gtt	137	7	1.393
8	cag	GTA	1505-1680	TCA	gtt	175	8	0.097
9	tag	GTG	1681-3715 (TAA at 1883)	/	/	2035	9	/

Table 8

Gene	Primer A	Primer B	#	T
Ex-1	CATGGCAACGGTTGGCTTG	GGCATGGTTCTTGTGAAAGC	1	55
Ex-2	ATCTGTCGCCAGGGTTCTAC	GTAACCTGAAGCAAAACAG	P	55
Ex-3	AGATGAAATTCTCAAGAACCTCTGAAATGGAGA	TTACGAGCTCCCTGGACTGTAAGCCTTG	P	55
Ex-4	ATGATCTACTGCTCCCACTG	CTCCACCAAGCACAAATAATG	P	55
Ex-5A	CTGATGCTGGCAAGAGGCT	CTTCCTGAGAACCTTGAAACAC	P	55
Ex-5B	ATTGTTCTTCCCCAACCTC	CTCCCCCTTCAGGATCCTG	P	55
Ex-6	CTTCCTTAACATGAGGAAC	CCAGTGGTCCAAACTAGTG	P	55
Ex-7	GTAGGTAGTTTGTCTTGT	CATCTAGGAAGGGAGAACT	P	55
Ex-7B	GGTTACTGTATGTCAGGAGA	TGCTTTGCCUTTCCCATTG	P	55
Ex-8	TTTGCAGTGAAGGAATGG	CAAAGAAGTCCACGTGTTG	P	55
Ex-9	CTTCCTATGAGAACAACTGG	CAAAGTCACTAGGATGTAG	P	55

GPC3	Primer A	Primer B	#	T
Ex-1	TGGGTAGGACCCCTCTT	CCTGGCACGGGAGCTTCCTGG	1	55
Ex-2	GTCCTCCCTAAGGGCCAAAC	CTTGGAAAACCGAGGACATC	2	55
Ex-3	AGAGGCCCTTGAAATTGTG	AGTGGTGGTCAGCTTCTG	2	55
Ex-4	ATTGCAAGTTATGTGCC	CGGCTGATAAGCTTCTTC	4	55
Ex-5	GAAGGAACTAATTAGAAG	TCTCCAGGAGTTCTGTCCA	1	55
Ex-6	ATAGAGCCAAAAGGAGCIA	ACCTCATGGAGATTGAAC	1	55
Ex-7	GATAGGAAAGGTGAAAG	ATCATTCATCACCAGAG	1	55
Ex-8A	ATCTGGATGTGGATGATG	AAAGGTGCTTATCTCGTTC	P	55
Ex-8B	ATCTGGATGTGGATGATG	GGCTGGGGAGGTATACGG	1	55

# : Stratagene Optiprime buffer (number) ; Perkin Elmer Buffer (P)

T : extension temperature

Table 9

GPC#	Primer A	Primer B	#	T
Ex-1	TCCCGTCCGCTCCAAAGGT	GGCATGGTCTCTGGTAAGC	6	55
Ex-2	TTCATGAATGATGATAAACATTGAC	CAGACCTAACTGCCTTCT	P	55
Ex-3A	CACTGAAAACCTCTGTTTTC	GGAGGCCAGAACCTCAATTA	P	55
Ex-3B	TACCTGGTGGAAATGCTGA	AAAAACAAACAGAGAACAGGA	P	55
Ex-4	TCTGTCCTCTATTAGCCCCG	CTTCAGCHAGCACAAATATG	P	55
Ex-5	ATTCCTCTCTCCAAACCTC	CTCCCTTTCAGCATCTG	P	58
Ex-6	CTTCCTAACATGGCACACTT	TGTCATTAACCGCCCTT	P	55
Ex-7	GTTAGTAGTTTGCTTGTC	CATCTAGGAAAGGGAAACT	P	58
Ex-8	CGAAATCACTTACTGGGTCA	CGTACTTCAAACCAAAATGC	P	55
Ex-9A	CTTCTGATGAGAAAGCTGG	AGAGAGAGTGGAGGGTAG	P	58
Ex-9B	AATGAGAAAGCCGCACTGC	CAAACCTACTAGATGTTAG	P	58

# : Stratagene Optiprime buffer (number) ; Perkin Elmer Buffer (P)

T : extension temperature

Table 10

	Patient ID	0817	0857	0893	0946	0969	0367*	13084	pCME
<b>GPC-3 Exon</b>									
1	+	+	+	+	+	+	+	+	-
2	+	+	+	+	+	+	+	+	-
3	+	(a)	+	(a)	+	(b)	+	+	+
4	+	+	+	+	+	+	+	+	+
5	+	+	+	+	+	+	+	+	+
6	+	+	+	+	+	+	+	+	+
7	+	+	+	-	+	+	+	+	+
8	+	+	+	-	+	+	+	+	+
<b>GPC-4 Exon</b>									
1	+	+	+	-	+	+	+	+	+
2	+	+	+	-	+	+	+	+	+
3	+	+	+	-	+	+	+	+	+
4	+	+	+	-	+	+	+	+	+
5	+	+	+	-	+	+	+	+	+
6	+	+	+	-	+	+	+	+	+
7	+	+	+	(c)	-	+	+	+	+
8	+	+	+	(d)	-	+	+	+	+
9	+	+	+	-	+	+	+	+	+

Table 11

GPc5	Left primer	Right primer
Exon -1	GGATGGACCAAGCACAGACCT	CCTGACAGGTCCACAGCG
Exon -2	TTCGATATTCAAAGGCCATCA	CCTGAAAAGCAGCCGCATT
Exon -3	GAACCCCTGAAACUCTCATCAA	GAGGTAGGCCCTAAACAGGATCATTA
Exon -4	CAAGTAAATAGGATTGTGG	GTCTCTTCACGTGTTCTTGTTGGTCT
Exon -5	GAGAAAGAATTTATCACAGCCCT	ATATCCTCTCCATTCAGCAGG
Exon -6	TATPACTCAGCTGTTGGTGG	ACAATCAGAACAACTATGTTTC
Exon -7	GTGATCACCACCTGAC	GGCATCOAGTCTGTAATCTTC
Exon -8	TCCATCAAACAGGACACTGGC	CATTAAACCTGAGTTAGCGA

Table 12

GPc6	Left primer	Right primer
Exon -1	GTTTACCGAGCTGGATTTG	AGCTCGAGCCCTTCACATC
Exon -2	TCTCTAGGAATAATCATGTC	TTCGTCAAATTCTTATGCTTG
Exon -3	AGTCCCTGAGATGCGAGAA	ACGTCCTCAATATGGCTTGAG
Exon -4	ATCCGTGCCCTCATGAAGTGC	AACAGATTCCACTCTGTGTG
Exon -5	GATGCAATGCTCTGGGG	ACCTGCAATGCTGTTCTCTG
Exon -6	CTTTCAGGGATGSGTCAGC	GTCCAAAGCTGTGCTGC
Exon -7	CAGACATAAAAGGAAATGAAAGC	GCTTTCCTGTECCGCTTC
Exon -8	GGTTCACCAACCGATCAT	GCCATTAATCTGCTGTCATCAAAGTC
Exon -9	GTTTCCACGGAAATTGAGT	AAGTGAAGTGGCCATTAGG

## REFERENCES

1. David et al., *J. Cell Biol.* **111**:3165-3176 (1990)
2. David et al., *J. Cell Biol.* **119**:961-975 (1992)
3. Filmus et al., *Biochem. J.* **311**:561-565 (1995)
4. Higgins, *Methods Mol. Biol.* **25**:307-318 (1994)
5. Matthijs et al., *Nat. Genet.* **16**:88-92 (1997)
6. Nakato et al., *Drosophila Development* **121**:3687-3702 (1995)
7. Pilia et al., *Nat. Genet.* **12**:241-247 (1996)
8. Steinfeld et al., *J. Cell Biol.* **133**:405-416 (1997)
9. Stipp et al., *J. Cell Biol.* **124**:149-160 (1994)
10. Vermeesch et al., *Genomics* **25**:327-329 (1995)
11. Veugelers et al., *Genomics* **40**:24-30 (1997)
12. Watanabe et al., *J. Cell Biol.* **130**:1207-1218 (1995)
13. Wiegant et al., *Genomics* **10**:345-368 (1991)
14. Zhang et al., *J. Biol. Chem.* **270**:27127-27135 (1995)

## CLAIMS

1. Polynucleotide encoding a glypcan-related protein, identified herein as "glypcan-6" and comprising at least the coding sequence of the nucleotide sequence as depicted in figure 1.
- 5 2. Glypcan-6 gene (GPC6) in isolated form comprising at least as the coding sequence the nucleotide sequence as depicted in figure 1 optionally interrupted by one or more introns, and optionally operably linked to transcription and translation regulatory sequences.
- 10 3. Polynucleotide encoding a glypcan-related protein, identified herein as "glypcan-4" and comprising at least the coding sequence of the nucleotide sequence as depicted in figure 2.
- 15 4. Glypcan-4 gene (GPC4) in isolated form comprising at least as the coding sequence the nucleotide sequence as depicted in figure 2 optionally interrupted by one or more introns, and optionally operably linked to transcription and translation regulatory sequences.
- 20 5. Derivatives of the polynucleotide sequence as depicted in figures 1 or 2, which derivatives are selected from fragments of the gene as claimed in claim 2 or 4, either isolated or synthetic and having a length that is smaller than the complete gene; primers, comprising at least 10 consecutive gene specific
- 25 nucleotides, preferably about 20 gene specific consecutive nucleotides of the nucleotide sequence of the gene; longer oligonucleotides up to the full length of the gene; antisense variants of the gene, the fragments or the primers; antibodies directed to the gene, fragments, primers or complementary strands thereof; any specific ligand for DNA that can be used as a specific probe, peptide nucleic acid probes.
- 30 6. Derivatives as claimed in claim 5, which derivatives are selected from transcripts (mRNA sequences) of the gene, cDNA, antisense RNA, antisense cDNA, antibodies directed to the transcript, sense and

antisense cDNA, antisense RNA and any specific ligand for RNA that can be used as a specific probe.

7. Derivatives as claimed in claim 5, which derivatives comprise at least part of the amino acid sequence encoded by the coding sequence of the nucleotide sequence depicted in figure 1 or 2 and selected from the isolated or synthetic gene product (protein or polypeptide); isolated or synthetic peptides, comprising a specific sequence of consecutive amino acids encoded by 10 the gene, antibodies directed to the gene product or peptides and any specific ligand for peptides that can be used as a specific probe.

8. Polynucleotides of claim 1 and 3, for use in diagnosis and/or therapy.

15 9. Gene of claim 2 and 4 and/or derivatives of claims 5-7 for use in diagnosis and/or therapy.

10. Method for diagnosing aberrations in a glypcan encoding gene, comprising isolation of the gene from cells expected to be harboring an aberrant gene; and 20 comparing the nucleotide sequence of the gene thus obtained with the nucleotide sequence of a wild type gene.

11. Method as claimed in claim 10, wherein 25 comparing the nucleotide sequence of the gene to be diagnosed with the nucleotide sequence of a wildtype gene is performed by restriction fragment length polymorphism screening, comprising separately digesting the isolated gene and a wild type comparison gene with one or more selected restriction enzymes, separating the digest thus 30 obtained on a gel to reveal a pattern of bands, and comparing the patterns of the isolated gene and the wildtype gene.

12. Method as claimed in claim 10, wherein 35 comparing the nucleotide sequence of the gene to be diagnosed with the nucleotide sequence of a wildtype gene is performed by means of Polymerase Chain Reaction (PCR), between probes corresponding to various parts of the gene to be diagnosed, for example exons, separating the

reaction mixture thus obtained on a gel to reveal a pattern of bands, and comparing the patterns of the isolated gene and the wildtype gene.

13. Method as claimed in claim 10, wherein  
5 comparing the nucleotide sequence of the gene to be diagnosed with the nucleotide sequence of a wildtype gene is performed by single-strand conformation polymorphism screening.

14. Method as claimed in claim 10, wherein  
10 comparing the nucleotide sequence of the gene to be diagnosed with the nucleotide sequence of a wildtype gene is performed by DNA sequencing.

15. Method for the in situ detection of physical changes in a glypcan gene, like translocations, 15 inversions or deletions, by the in situ hybridization of labeled probes with a set of chromosomes.

16. Method as claimed in claim 15, wherein translocations can be detected by hybridizing a set of chromosomes with a first probe that hybridizes to a part 20 of the glypcan gene that is not likely to be involved in the translocation, and a second probe that hybridizes to a part of the glypcan gene that is likely to be involved in such aberration, wherein translocations are identified when the second probe is detected on another chromosome 25 than the first probe.

17. Method as claimed in claim 16, further comprising identification of the translocation partner, and using in addition probes hybridizing to the translocated part and the remaining part of the 30 translocation partner and bearing a different label from the first set of probes.

18. Method as claimed in claim 15, wherein an inversion is identified if the second probe is found closer to or further away from the first probe than in a 35 non-aberrant chromosome.

19. Method as claimed in claim 15, wherein deletions are detected when one or both of the probes are not present on the aberrant chromosome.

20. Method as claimed in claims 15-19, wherein the gene to be diagnosed is the glypican 3 gene and the probes are as given in Table 10.

21. Method as claimed in claims 15-19, wherein 5 the gene to be diagnosed is the glypican 4 gene and the probes are as given in Table 10 and/or 11.

22. Method as claimed in claims 15-19, wherein the gene to be diagnosed is the glypican 5 gene and the probes are as given in Table 13.

10 23. Method as claimed in claims 15-19, wherein the gene to be diagnosed is the glypican 6 gene and the probes are as given in Table 14.

24. Method as claimed in claims 15-19, wherein the gene to be diagnosed is the glypican 1 gene and the 15 probes are derivable from figure 3.

25. Method for diagnosing the expression pattern of glypican genes, wherein antibodies directed to the gene product or protein are reacted with Western blots of cell extracts to detect the presence of the 20 protein in the cell or to assess the amount of protein present.

CTCTCACAGGCTCCACAGGCCACAGCACACATCCCCAAGAACCTCGAGCTCACACCAACAGACAC  
 CCAACACATAACCTCAGGCCACACACTCGCTTCGGTCACTGGCTTGTCCATTCTCCGGGGGAGCCGGCG  
 ACGCGGCATACACACTCGCTTCGGTCACTGGCTTGTCCATTCTCCGGGGGAGCCGGCG  
 CTCACCTGGATTCGGCATACACACTCGCTTCGGTCACTGGCTTGTCCATTCTCCGGGGGAGCCGGCG  
 TCGGAACCTCGGATTCGGATTCGGTCACTGGCTTGTCCATTCTCCGGGGGAGCCGGCG  
 CTCTCCCTCGTTGATTGACCCGTTTCATCTGGGGCTAGAGGAGCAAGGAGCAGCC  
 TTCCAGGCCAGCCCTGGCTTGCCATCGTCCATCTGGCTTATAAAAGTTGCTGAGC  
 GCAGTCCAAGGGCTGCGCTGCTGGCTTCTGGGGCTTCTGGCTTGAAGGGCAAGGTGAAGAG  
 GGGCGAGGAGGCCGGCTGCTGGCTTACCGAGCTGGATTTGATGTCGACCATTGCTTGGATC  
 CGCACCGCCGTTGGGTTACCGAGCTGGATTTGATGTCGACCATTGCTTGGATC  
 M P S W I 1/33  
 60 120 180 240 300 360 420 480 540 600  
 660 25 720 45 780 65 840 85 900 105 960 125 1020 145  
 G A V I L P L L S L P A G A D V  
 PAGGCTCGGAGGGTCCGGAGGGCTGGCTGGGTACGGCTGCCCCTCCCCGGGGGGATG TG  
 K A R S C G E V R Q O A Y G A K G F S L A  
 GACATCCCCTACCAAGGAGATGCCAGGGAAACTTAAGAATCTGTCTCAGGAATATAACA  
 D I P Y Q E I A G E H L R I C P Q E Y T  
 TGCTGCACCAAGAACATGGAAGAACAGTTAACGCCAACAAAGCCTAACACTGAAATTGAAAAC  
 C C T T E M E D K L S Q Q S K L E F E N  
 CTTGTGGAAGGACAAGGCCATTGTGCGCACCATTGTGTCAGGGCATAAAGAAATT  
 L V E E T S H F V R T T F V S R H K K F  
 GACGAATTTCGGAGAGCTCCGGAGAATGCGAGAAAAGTCACTAAATGATAATGTTGTA  
 D E F F R E L L E N A E K S L N D M F V  
 CGGACCTATGGCATGCTGTACATGCCAGAATTCCAGAATGCTTCCAGGACCTCTCACAGAG  
 R T Y G M L Y M Q N S E V F O D I F T E

CTGAAAAGGTACTACACTGGGGTAATGTGAATCTGGAGGAATGCTCAATGACTTTGG  
 L K R Y T G G N V N L E E M L N D F W 1080  
 GCTGGCTCCTGGAACGGATGTTCAAGCTGATAAACCCTCAAGTATCACTTCAGTGAAGAC 1165  
 A R L L E R M F Q L I N P Q Y H F S E D 1140  
 TACCTGGAAATGTGTGAGCAAATAACACTGACCAGGTCAAGCCATTGGAGACGTGCCCGG 185  
 Y L E C V S K Y T D Q L K P F G D V P R 1200  
 AAACGTAAAGATTCAAGGTTACCCGGCGCCTCATGGCTGCCAGGGACCTTGTCCAGGGGCTG 205  
 K L K I Q V T R A F I A A R T F V O G L 1260  
 ACTGTGGGCAGAGAAGTGGCAAACCGAGTTCCAAAGGTCAAGCCCAACCCCCAGGGTGTATC 225  
 T T V G R E V A N R V S K V S P T P G C I 1320  
~~G~~<sup>2/33</sup> CGTGCCTCATGGAAAGATGCTGTACTGCCCATACTGTGGGGCTTCCCAACTGTGAGGGCC 245  
 R A L M K M L Y C P Y C R G L P T V R P 1380  
 TGCAACAACTACTGTCTCACGTCATGAAGGGCTGCTGGGAAATCAGGCTGACCTCGAC 265  
 N C N N Y C L N V M K G C L A N Q A D L D 1440  
 ACAGAGTGGAAATCTGTTTATAGATGCAATGCTCTGGTGGCAGGGCAGCTGGAGGGCCA 285  
 T E W N L F I D A M L L V A E R L E G P 1500  
 TTCAACATTGAGTCGGTCATGGACCCGATAGATGTCAAGATTCTGAAGCCATTATGAAAC 305  
 F N I E S V M D P I D V K I S E A I M N 1560  
 ATGCAAAACAGCATGCAAGGTGTCTGCAAAGGTCTTCAGGGATGTTGGTCAGCCCCAA 325  
 M Q E N S M Q V S A K V F Q G C G Q P K 325  
 CCTGCTCCAGCCCTCAGATCTGCCCGCTCAAGCTCTGAAATTTAAATACACGTTTCAGG 345  
 P A P A L R S A R S A P E N F N T R F R 1620  
 CCCTACAAATCTGAGGAAAGACCAACAACTGCTGGCAGGCCAACAGCTGGTGGCTGG 1680  
 P Y N P E E R P T T A A G T S L D R L V 1740  
 ACAGACATAAAAGAGAAATTGAAGCTCTCTAAAAAGGTCTGGTCAGGCATTACCCCTACACT 385  
 1800

T D I K E K L K L S K K V W S A L P Y T  
 ATCTGAAAGGACGGAGCGTGCACGGGGCACGTCCAACGAGGGAGAATGGCTGGAACGGG  
 I C K D E S V T A G T S N E E C W N G 425  
 CACAGCAAAGCCAGATACTTGCTTGAGGATCATGAATGATGGCTCACCAACCAGATCAAT  
 H S K A R Y L P E I M N D G L T N Q I N 445  
 AATCCCAGGGTGGATGGACATCCTGGCCTGACACTTTCATCAGACAGGAGATTATG  
 N P E V D V D I T R P D T F I R Q Q I M 465  
 GCTCTCGTGTGATGACCAACAAACTAAAAACGCCCTACAATGGCAATGATGTCATTTTC  
 A L R V M T N K L K N A Y N G N D V N F 485  
 CAGGACACAAAGTGAATCAGTGGCTCAGGGGAGTGGCAGTGGATGACGTTG  
~~T~~<sup>Q</sup> D T S D E S S G S G S G C M D D V 505/33  
~~G~~<sup>T</sup> GTCCCCACGGAGTTGAGTTGTCACCAACAGGGCCCCCGCAGTGGATCCGACCGGAGA  
~~T~~<sup>C</sup> P T E F E V T T E A P A V D P D R R 525  
~~U~~<sup>E</sup> JGAGGTGGACTCTCTGCAGGCCACTCCGTGCTCTGGTCTCTCAGCTGC  
~~E~~<sup>V</sup> D S S A A Q R G H S L L S W S L T C 545  
 ATTGGTCTGGCACTGCGAGAGACTGTGCZAGATAATCTTGGGTTTTGGTCAGATGAACTG  
 I V L A L Q R L C R \* 555  
 CATTTTAGCTATCCTGAATGGCCAACCTCACTTCTTCTTACACTCTTGACAATGGACC  
 ATGCCACAAAACCTTACCGTTTCTATGAGAAGAGAGCAGTAATGCAATCTGCCTCCCTT  
~~T~~<sup>TT</sup> TTGTTTCCCAAAGAGTACCGGGTGGCAGACTGAAGCTGCTTCTCTTCAGCTAT  
 CTGTGGGGACCTTGTGTTATCTAGAGAAATTCTTACTCAAAATTCTGTTACCGGAGAT  
 TTTCTTACCTTCATTGCTTTATGCTGGAGAAGTAAGGAATCTACGTTGTGAGGGTT  
 TTTTTTTCTCATTTAAATAAAAGGAAGAAAGAAAATAATTCTCTGTAAATCTG  
 GCCAAACCCCAAGACAGCTACATTTCACAAACAAAAAGCAAACAGAGAAAATATGAA  
 CTTAACACTGTAAGTTGACATGGCC 2700  
 2731

4/33

GCCTGGCACCGGGACCGTTGCCTGACGGAGGCCAGCTACTTTGCCCCCGGTCT  
 CCTCCGGCTGGTCGGCCTCTTCCACCAACTCCAACTCCTTCCAGCTCCACTCGCT  
 AGTCCCCGACTCGCCAGGCCCTGGCCGCTGGCGTAGGCCGTTCCCGTCCGGTCCA  
 AAGGTGGGAACGTGTCCGGCCGGGGCACCATGGACGGTTGGCTGCGAGCTCAAGTGAAAAGTT  
 TCTGCACCCCTGGCAGTGTCAAGGCCGGCTGGCTGGCTGCCAGCTCAAGTGAAAAGTT  
 C T L A V L S A A L L A A E L K S K S C  
 GCTCGGAAGTGGAGCTTACGTTGCTCAAAGGGTTCAACAAGAACGATGCCCTCC  
 S E V R R L Y V S K G F N K N D A P L H  
 ACGAGATCAACGGGTGATCATTTGAAGATCTGTCCCCAGGGTTCTACCTGCTGCTCAAG  
 E I N G D H L K I C P Q G S T C C S Q E  
 AGATGGAGGAGAAGTACGCCCTGCAAAGTAAGATGATTICAAAAGTGTGGTCAGCGAAC  
 M E E K Y S L Q S K D D F K S V V S E Q  
 AGTGCAATCATTTGCAAGGTGTCTTGTCTCACCGTTACAAGAAAGTTGATGAATTCTTCA  
 C N H L Q A V F A S R Y K K F D E F F K  
 AAGAACTACTGAAAATGGAGAAAATCCCTGATGATGATGTTGTGAAGACATATGGCC  
 E L L E N A E K S L N D M F V K T Y G H  
 ATTATACATGCAAAATTCTGAGCTTAAAGATCTCTTGTAGAGTTGAAACGTTACT  
 L Y M Q N S E L F K D L F V E L K R Y Y  
 ACGTGGTGGAAATGTGAACCTGGAAAGAAATGCTAAATGACTTCTGGCTCGCCTCCTGG  
 V V G N V N L E E M L N D F W A R L L E  
 AGCGGATGGTCCGGCTGGTGAACCTCCAGTACCAACTTACAGATGAGTATCTGGAAATGTC  
 R M F R L V N S Q Y H F T D E Y L E C V  
 TGAGCAAGTATACGGAGCAGCTGAAGCCCTTCGAGATGTCCTCGCAAATTGAAGCTCC

-1

S K Y T E Q L K P F G D V P R K L K L Q  
 AGGTACTCGTGCCTTTGCTAGCAGCCGTACTTCGCTCAAGGCTTACGGGTTGGGGAG 210  
 V T R A F V A R T F A Q G L A V A G D 900  
 ATGTCGTGAGCAAGGTCCTCCGTGTTAACCCCCAACAGCCCAGTGTACCCATGCCCTGTGA 230  
 V V S K V S V N P T A Q C T H A L L K 960  
 AGATGATCTACTGCTCCACTGCCGGGGTCTCGTGAAGGCCATGTTACAACTA 1020  
 M I Y C S H C R G L V T V K P C Y N Y C 250  
 GCTCAAACATCATGAGAGGCTGTTGGCCAACCAAGGGGATCTCGATTGTTGAATGGAAACA 270  
 S N I M R G C L A N Q G D L D F E W N N 1080  
~~F~~  
 ATTTCATAGATGCTATGCTGATGGCTGATGGTGGCAGAGGGCTAGAGGGTCCCTTCAACATTGAAT 290  
 F I D A M L M V A E R L E G P F N I E S 1140  
 CGGTICATGGATCCCACATCGATGTTGAAAGATTCTGATGCTTATTATGAAACATGCAGGATAATA 310  
 V M D P I D V K I S D A I M N M Q D N S 5/33  
 GTGTTCAAGTGTCTCAGAAGGTTCCAGGGATGTGGACCCCCAAGGCCCTCCAGCTG 1200  
 V Q V S Q K V F Q G C G P P K P L P A G 330  
 GACGAAATTCTCGTTCATCTCTGAAAGTGCCTTCAGTGCCTCAGACCACATCACC 1260  
 R I S R S I S E S A F S A R F R P H H P 350  
 CCGAGGAACGCCAACCCACAGCAGCTGGCACTAGTTGGACCGACTGGTTACTGTATGTCA 1320  
 E E R P T T A A G T S L D R L V T D V K 370  
 AGGATAAACTGAAACAGGCCAACAGAAATTCTGGTCTCCCTCCGAGCAACGTTGCAACG 1380  
 D K L K Q A K F W S S L P S N V C N D 410  
 ATGAGGAGATGGCTGCAGGAAACGGCAATGAGGATGACTGTTGGAAATGGGAAAGGCAAA 1500  
 E R M A A G N G N E D D C W N G K G K S 430  
 GCAGGTACCTGTTGCACTGACAGGAAATGGATTAGTCACCAGGGCAACACCCAGAGG 1560

2-2

R	Y	L	F	A	V	T	G	N	G	L	V	N	Q	G	N	N	P	E	V	450
TCCAGGTTGACACCAGCAAACCAGACATACTGATCCTTCGTCAAATCATGGCTCTTCGAG																				1620
Q	V	D	T	S	K	P	D	I	L	I	L	R	Q	I	M	A	L	R	V	470
TGATGACCAGCAAGATGAAAGATGCCATAATGGGAACGACGTGGACTCTTGTGATATCA																				1680
M	T	S	K	M	K	N	A	Y	N	G	N	D	V	D	F	F	D	I	S	490
GTGATGAAAGTAGTGGAGAAGGAAAGTGGGAAGTGGCTGTGGAGTATCAGCAGTGCCCTTCAG																				1740
D	E	S	S	G	E	G	S	G	C	E	Y	Q	Q	C	P	S	E	510		
AGTTTGAECTAACATGCCACTGACCATGCTGGGAAGAGTGCCAATGAGAAAAGCCGACAGTG																			1800	
F	D	Y	N	A	T	D	H	A	G	K	S	A	N	E	K	A	D	S	A	530
CTGGTGTCCGTCTGGGCACAGGCCAACCTCTACACTGCTCTACTGCTCTGCATCTTGTCTGG																				1860
G	V	R	P	G	A	Q	A	Y	L	T	V	F	C	I	L	F	L	V	550	
TTATGCAGAGAGAGTGGAGATAATTCTCAAACCTGTGAGAAAAGTGTTCATCAAAAGTT																				1920
M	Q	R	E	W	R	*														556
AAAAGGCACCAGTTATCACTTTCTACCATCCTAGTGTCTTTTAATGAATGGA																				1980
CAACAATGTAACAGTTTACTATGTTGCCACTGTTGTTAACGAAAGTGCTGACTTTGTTCT																				2040
CATTCACTGTTGGAGAAAGGGACTGTGCAATTGAGTTGGTCTCTGCTCCCCAAACCA																				2100
TGTTAAACGTGGCTAACAGTGTTAGGTACAGAACATACTAGTTAGTTGTCATTGGATT																				2160
ATCAGCTCTATTATTTGTTGTATGTTTTCATTTCGTTGGGTTTTTTTCCA																				2220
ACTGTGATCTCGCCTTGTCTTACAAGCAACCCAGGTCCCTCTGGCACGIAACATG																				2280
TACGTATTCTGAAATTAATAGCTGTACAGAAGCAGGTTTATTATCATGTTATCT																				2340
TATTAAGAAAAAGCCAAAAAGCAGTAAATTCCATTCTCCCTGTTATTATTAAGACAGAATG																				2400
CCTTATCTGGAGAGACGTGGAGGTGATTTCCTTTTAAATTATTAAGACAGAATG																				2460

FIG. 2-3

FIG. 2 - 4

TGAAGGCCAAGCAGGCTTCTGAGCCACTTGTAGATTGATTCAGATTCAATCCAAG	2520
AGGAGTTATGGTACTTCATTATTGGTGATAAGTGAGACTGAGACTACTGCTT	2580
TGAATGAGTTGAATTACATAAGCTAAGATCACTATAAGGCCATTCTTGAAACCCACTTA	2640
TACATAAAATGTAACCCATTAGAAAAGATTCTGGGATATCATCCCCCCTTGAAGATAG	2700
AAAGCATTCAAGATGGTCCCACTTACATGTTCACACTGGGTTAGGGGTGTTTTTT	2760
TTAAACCAGGCAGGTTAGCTAGCCACCCCTGTGCTAGTTTCATGTTACGGCTGACCCCT	2820
ATTGAAATTAAATACCTTGTAGTGGTGCAGATTCAACCCAAATTATGACANGGA	2880
GCTGCTGAGAGGCTTAGGCCAGAACCTGGGGTACAGGCCTGGGCTCAGGGATAAGCTGTCAACA	2940
CTCGGGCAAAAGTTTTGGCTGGCAGCAACGGTGTATCACCATTGTTGGGATCCAGTTT	3000
TGTTTAAGAGAGTATAAGGGTGTCTCATTGATGCTTTTACCTAGCCCCCTTTAT	3060
CAGTAAACAAAGGACTTGCCATGGTCAACGCAATGGCTACGATCCAAGATACTCAACC	3120
AAGGAGCCCACCTAGGGGAGAACCTAGGGTGTCCAGATTGTTGATGTTCTTGTGG	3180
AGGGATGGGGTGGGGTGGGAGTAGGTAGAGCTGAGAAATACATACATCTAGTGGTGCCTT	3240
AGCCACGTGGGTGAAGTGGCAAAGGCCATGGCATAATCTAGTGGTGCCTT	3300
ACAACCTGCCCTGGGAATCCCTTCCCTGTGTCCTTACCAAAATGATAGCTCATAAAACCTG	3360
ATAATGTAACAAATCACTTNCAAGGAGTTCCAGAAGTCTCAGAAAGACTAAATTCT	3420
GTCTCTCCCTGCTTAGACAGGCCATTAAAGATCCAACATAATTACCGAACCTAAACCC	3480
ACAAAGAGGTTGGTTGTTGTTATTGTTCAATCTCAGTTGTAAGAGTAATTCTCTATT	3540
ATATGAAACATAATTACTTGATAGCTCAGGGCTACACTTCATTCAACTTITACCCA	3600
AATTCTGCAGAGTGGTCAAAATGGAAATTGGGGCTGTGTGAAACAGGGCTTAATT	3660
ATTAGAAGTAGGCCAGTTATTAAAGCATGATGTAAATAGGCATATTCCAAA	3720
AAAA	3724

8/33

A R L E R L F K Q L H P Q L I L P D D  
 ACTACCTGGACTGGCCTGGCCAAAGGAGGGCGAGGCCCTTGGGGAGGGCCCCGA 840  
 Y L D C L G K Q A E A L R P F G E A P R 207  
 GAGAGGCTGGCCTGGGGCCACCCGGTCGTGGCTGCCTCTTGTGCAAGGGCC 900  
 E L R L R A T R A F V A A R S F V Q G L 227  
 TGGGGTGGCCAGCGAACGGTGGCTACTGTGCTACTGCTGGAGTCCCCGGCCAGGC 960  
 G V A S D V V R K V A Q V P L G P E C S 247  
 CGAGAGCTGTATGAAGCTGGTCTACTGTGCTACTGCTGGAGTCCCCGGCCAGGC 1020  
 R A V M K L V Y C A H C L G V P G A R P 267  
 CCTGGCCCTGACTATTGGCGAAATATGGCTCAAGGGGCTGCCAACCGAGGCCACCTGG 1080  
 C P D Y C R N V L K G C L A N Q A D L D 287  
 ACGCCGAGTGGAGGAACCTCCTGGACTCCATGGTGTCAATCACCGACAAAGTTCTGGGTA 1140  
 A E W R N L L D S M V L I T D K F W G T 307  
 CATCGGGTGGAGAGTGTCAATGGGAGCGGTGGCACAGTGGGGAGGCCATCAACG 1200  
 S G V E S V I G S V H T W L A E A I N A 327  
 CCCTCCAGGACAACAGGGACACGGCTACGGCTAACGGTCATCCAGGGCTGGGGAAACCCA 1260  
 L Q D N R D T L T A K V I Q G C G N P K 347  
 AGGTCAACCCCCAGGGCCCTGGGCCATGGAGAAGGGCAAGCTGGGGCAAGCTGGCCCGC 1320  
 V N P Q G P G P E E K R R G K L A P R 367  
 GGGAGGGCCACCTTCAGGGCAGGTGGAGAAGGTGGCTCTGAAGGCAAGGCCAGCTCC 1380  
 E R P P S G T L E K L V S E A K A Q L R 387  
 GCGACGTCCAGGACTTCTGGATCACGGCTCCAGGGACACTGTGGCAGTGGAGAAGTGGCC 1440  
 D V Q D F W I S L P G T L C S E K M A L 407  
 TGAGGACTGCCCAGTGTATGACCGCTGGTGAACGGGATGGCCAGAGGCCGTACCTCCCCG 1500

3-2



11/33

2580 GGAGTCTGAGGACTGTCCCTCCCACAGACCCCTGCAGTGAGGGCCCTCATGGCGAGATGA  
 2640 GGGGCCACTGACCCACCTGCGCTTCGCTGGGAACTGGGCAATGTTGGCTGCCACACAGGGCTCACAGGGCAG  
 2700 GCCTTGCTGGGTCCAGGGCTGTTGAGGACCCCGAGGGCTGAGGGCAGGCCAGGACCCG  
 2760 CCTGCTCCATCCTACCCAGATCAGGAACCAGGGCTCCCTGTTCAAGGTGACACAGGT  
 2820 CAGGGCTCAGAGTGAACCTCGGGCTGACCTGTCACAGGGATGCTGGCTGGTGGTGA  
 2880 11/33  
 2940 CCCCGCACTGCAACACGGGAATGGCTTCCCGAACCCAGGCCAGCTGCAGCTGCAGG  
 3000 3060 3120 3180 3240 3300 3360 3420 3480 3540 3600 3660 3692  
 3060 3120 3180 3240 3300 3360 3420 3480 3540 3600 3660 3692  
 GCACGGGACCTGGATAGTTAAGGGCTTTTCCAAACATGCACTCCATTACTGACACTTCC  
 TGTCTTGTTCATGGAGAGCTGTTGCTCTCCAGATGGCTCGGAGGCCAGGGCC  
 CACCTTGGACCCCTGGTGAACCTCCCTCACTGACTGAGGCATCAGGGCCCTGGCC  
 CTGGACGGGCCCTCCTCCTGCCCCAGCTGCCAGGTGGCCCTGGGAGGGTGG  
 TGTGGTGTGGGAAGGGGCTCTGCAGGGGAGGAGACTGGGGCTGGGGCT  
 TCCTGAACCGAATGACCCCTGAGGGGGCGCTTAAGTGTGCTGGTTTGCTT  
 CGGACAGTGGACGGAGGTCCCCGGTTGCTGGTCAGGTCCCCATGGCTT  
 CTGACTTTAGATGTTGGATCAGGAGCCCCAACACAGGAAGTCCACCCZATAATA  
 CCCTGCCAGTGCCAGGGTGGGACTTGGGACTCTGGCACAGTGATGCCGGCAG  
 CAGGACTCCCCGCTGCACACAGACGGCCTAGGGTGGGCTCAGACCCACCC  
 TCTCTGGAAAGGGCAGGCCCTGAGTGGTCACTGGTCAGGGCAGTGGCCA  
 CCTCTCCACAAGGCCCCACGGCTCAGTGTCAAGGGTGGTCA  
 TCCTGTATGAATAAGGCTGGAAACCTAAA

FIG. 3-4

12/33

60	120
GTCGGCACCGGGTGCCTGGGTGGATGCTCAGCTTGACTTCCGGAGCAGGAAGCTGGACTGCTCTC	M A G T
GCTCCGGCTGCCACTCTCCGGCTCCTAGCTCCCTGGGAAGCAGGATGGCCGGAC	4
V R T A C L V V A M L L S I D F P G Q A	24
CAGCCCCGGCCGGACGCCACCTGTCAACCAAGTCCGCTCTTCCAGAGA	180
Q P P P D A T C H Q V R S F F Q R	240
CTGCAAGCCGGACTCAAGTGGGTGCCAGAAACTCCCGTGCAGGATCAAGTTGAAAGTA	44
L Q P G L K W V P E T P V P G S D L Q V	300
TGTCTCCCTAACGGCCCCAACATGCTGCTCAAGAAAGATGGAAAGAAAATAACCAACTAAACA	64
C L P K G P T C C S R K M E E K Y Q L T	360
GCACGGATTGAACATGGAACAGCTGGCTTCAGTCTGCAAGTATGGAGCTCAAGTTCTTAATT	84
F A R L N M E Q L L Q S A S M E L K F L I	420
- ATTCAAGATGCTGGGGTTTCCAAGAGGGCCTTGAATTGTTGTTGCCATGCCAAGAAC	104
I Q N A A V F Q E A F E I V V R H A K N	480
TACACCAATGCCATGTTCAAGAACACTACCCCAAGCTGACTCCACAAAGCTTTGAGTT	124
Y T N A M F K N N Y P S L T P Q A F E F	540
GTGGGTGAATTTCACAGATGTGTCTCATCTGGGTTCTGACATCAATGTAGAT	144
V G E F F T D V S L Y I L G S D I N V D	600
GACATGGTCAATTGAATTGTTGACAGCCCTGTTCCAGTCATCTATAACCCAGCTTAATGAAAC	164
D M V N E L F D S L F P V I Y T Q L M N	660
CCAGGGCTGCCTGATTCAAGCTGGACATCAATGAGTGCCTCCGAGGCAAGACGTGAC	184
P G L P D S A L D I N E C L R G A R R D	720
CTGAAAGTATTGGAAATTCCCCAACGTTTATGACCCAGGTTCAAGTCACTGCAA	204
L K V F G N F P K L I M T Q V S K S L Q	780
GTCACTAGGATCTTCCTTAGGCTCTGAATCTTGGAAATTGGAAACTGATCAACACAACTGAT	224
	840

FIG. 1



<p>K L K H I N Q L L R T M S M P K G R V L  <b>GATAAAACCTGGATGAGGAAGGGTTTGAAGTGGAGACTCGGGTGATGATGAAGATGAG</b>  D K N L D E E G F E S G D G D E D E  <b>TGCATTGGGGCTCGGTGATGGAATGATAAAGTAAGGAAATCAGCTCCGCTTCCCTTGCA</b>  C I G G S G D G M I K V K N Q L R F L A  <b>GAACTGGCCTATGATCTGGATGTGGATGATGCCCTGGAAACAGTCAGCAGGCCAACCTCCG</b>  E L A Y D L D V D A P G N S Q Q A T P  <b>AAGGACAAACGAGATAAGCACCTTTCAAAACCTCGGGAAACGTTCATCCCCGCTGAAGCTT</b>  K D N E I S T F H N L G N V H S P L K L  <b>CTCACCGCATGGCCATCTCGGTGGTCTCTCCTGGTGCACTGACTGCGCTGGT</b>  L T S M A I S V V C F F F L V H *  <b>CCCCAGCACATGGCTGCCCTACAGCACCCTGTTGGTCTTCCTCGATAAAGGGAAACCACTT</b>  <b>TCTTATTTTCTTATTTTTTGTATTCCTGTATACCTCTCAGCCATGAAGT</b>  <b>AGAGGACTAACCATGTTATGTTGAAATCAAATGGTATCTTTGGAGGAAGATACT</b>  <b>ATTTTAGTGGTACCATATAGATGTCCTTTGCAAAAGAAAAAACCATCAAGTTG</b>  <b>TGCCAAATTATTCTCCCTATGTTGGCTGCTAGAACATGGTACCATGTTCTCTCTCA</b>  <b>CTCCCCTCCCTTCTATCGTTCTCTCTGATGGATTCTTGAAGAAAATAATTGCT</b>  <b>CAAATAAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAGAAAAAAAG</b></p>	484 1620 504 1680 524 1740 544 1800 564 1860 580 1920 1980 2040 2100 2160 2220 2280 2300
--	--

CCAGGACGGCGAGGATGGCACAGACCTGGCCGGTGGCTTCGGCTCCCTCCTTC  
 M D A Q T W P V G F R C L L L L 60  
 TGGCCCTGGTTGGGTCCCGGAGCGAGGGGGTGCAGACTGCGAAAGAAGTGGAAAC  
 A L V G S A R S E G V Q T C E E V R K L 16  
 120  
 A L V G S A R S E G V Q T C E E V R K L 36  
 TTTTCAGTGGGGCTGTTGGAGCTGTCAGGGGGCTGCCGGATTGCCGGGGCAGGAC  
 F Q W R L L G A V R G L P D S P R A G P 180  
 180  
 CTGATCTCAGGTTGCATATCCAAAAGCCTACATGTTGCACCAAGGAAGATGGAGGA  
 D L Q V C I S K K P T C C T R K M E E R 240  
 240  
 76  
 G GATATCAGATTGGGCTGCCAGGATATGCAGCAGTTCTCAAACGTCAGCTCAT  
 Y Q I A R Q D M Q Q F L Q T S S T L 300  
 300  
 TAAAGTTCTAATATCTCGAAATGGGCTGCTTTCAAGAACCCCTGAAACTCTCATCA  
 K F L I S R N A A A F Q E T L E T L I K 96  
 360  
 AACAAAGCAGAAATTACACCGATACTCTTGCAGTACCTACAGGAACATGGCCTTGG  
 Q A E N Y T S I L F C S T Y R N M A L E 116  
 420  
 AGGCTGCTGCTCGGGTCAAGGAGTTCTTCACTGATGGGGCTGTATTATTGGTGGGG  
 A A A S V Q E F F T D V G L Y L F G A D 136  
 480  
 ATGTTAATCCTGAAATTGTAACAGATTGGTAAATTCGACAGTCTTCTGGTCTACA  
 V N P E E F V N R F F D S L F P L V Y N 156  
 540  
 ACCACCTCATTAACCCCTGGTGTGACTGACAGTCCCTGGAAATCTCAGAAATGCCGA  
 H L I N P G V T D S S L E Y S E C I R M 176  
 600  
 TGGCTCGCCGGATGTGAGGCCATTGGTAATTTCCCACAGTAATGGGACAGATGG  
 A R R D V S P F G N I P Q R V M G Q M G 196  
 660  
 GGAGGGTCCCTGCTGCCAGGGCACTTTCTGCAAGGCACCAATCTGGGCATTTGAAGTCA  
 216  
 720

R S L P S R T F L Q A L N L G I E V I 236  
 TCAACACCACAGACTATCTGGCACTTCTCCAAAGAGTGGCAGGCCCTCCTGAGATGC 780  
 N T T D Y L H F S K E C S R A L L K M Q 256  
 AATACTGCCCGCACTGCCAAGGCCCTGGGCTCACTAACGCCCTGTATGGATACTGCCCTCA 840  
 Y C P H C Q G L A L T K P C M G Y C L N 276  
 ATGTCATGGCAGGGCTGGCTGGCGCACATGGCGGAGCTTAATCCACACTGGCATGGCATATA 900  
 V M R G C L A H M A E L N P H W A Y I 296  
 TCCGGTGGTTGAAAGAACACTCTGGATGCAATGCAATGGAAACATACGACATTTGGACACGGTGC 960  
 R S L E E L S D A M H G T Y D I G H V L 316  
 TGCTGAACTTCACTTGCTTGTAAATGATGGCTGTGTTACAGGCTCACCTCAATGGACAAA 1020  
 L N F H L V L N D A V L Q A H L N G Q K 336  
 ATTATTGGAACAGGTAATAAGGATTGGCCGCCCTGTTAAGAACCCCCACACAAAGGCC 1080  
 L L E Q V N R I C G R P V R T P T Q S P 356  
 CCCGGTTGGTCTTTGATCAGAGCAAAGAGAACGGATGGAAATGAAGAACCAAGGAACA 1140  
 R C S F D Q S K E K H G M K T T R N S 376  
 GTGAAGAGACGCCCTGGCAACAGAACAGAAATACTAACAGGCCCTGGACTGTGACAGGT 1200  
 E E T L A N R R K E F I N S L R L Y R S 396  
 CATTCTATGGGGTCTAGCTGATCAGCTTGTGCTTAATGAAATTAGCTGCTGCAGATGGAC 1260  
 F Y G G L A D Q L C A N E L A A D G L 416  
 TTCCCTGGTGGAAATGGAGAAGATATAGTAAAGTTATACTCAGGTTGTTGGAAATG 1320  
 P C W N G E D I V K S Y T Q R V V G N G 436  
 GAATCAAAGCCAGTCTGGAATCCCTGAAGTCAAAGTCAAAGGAATTGATCCTGTGATAA 1380  
 I K A Q S G N P E V K V K G I D P V I N 456  
 ATCAGATTATTGATAAACTGAAGCATGTTGTTACAGGGTAGATCACCCAAAC 1440

FIG. 5-2

Q I I D K L K H V V Q L L Q G R S P K P 476  
 CTGACAACTGGGAACCTCTTCAGCTGGGCACTGGTGGGGCATGGTTGAAACAAGTCAGTG 1500  
 D K W E L L Q L G S G G M V E Q V S G 496  
 GGGACTGTGATGATGATGAAAGATEGGTTGGGGGGATCAGGAAGGGAGAAGTCAAGAGGACAC 1560  
 D C D D E D G C G S G S G E V K R T L 516  
 TGAAGATCACAGACTGGGATGCCAGATGATGAAACTTCAGTGTGATGTAAGCAAATCCATC 1620  
 K I T D W M P D M N F S D V K Q I H Q 536  
 AACAGAGACACTGGCAGTACTTTAGACACAAACAGGAGCAGGATGTGCAGTGGCGACTGAAAT 1680  
 T D T G S T L D T T G A G C A V A T E S 556  
T CTATGACATTCACTCTGATAAGTGTGTTGATGTACTTCCGGGATTGGTAACTGAACT 1740  
 M T F T L I S V M L L P G I W \* 572  
 CTTCTGCTGACATACCTTACTGAAAGTCTCGATTTCCTCTCTGCATATGCCCTGGA 1800  
 ATAAGAGATCCATTCAATGTAACATTATTATTTATGAAAAGATAATGTTACACTAACCT 1860  
 CTCAGAAGCCAAGCTGAAATTCTATAAAGTCCCTAAACCTCAACGTTAAATGZACAC 1920  
 TTTAAAATAATATGCTTTCAATCTAACTGAAACCTCTTAACCTCTAAATATTTAAA 1980  
 TCTGAAGATGTGAAGGGCACAGAACAGTGAAGTGAATAGAAGAATTAGTGTATCTGAA 2040  
 TTTTATATCAATTCCAAGGCCCTTCTCTCTCTAAATTAAAAAATGTTCAATTGAAAGTG 2100  
 TATTGCCAGACAATGAAAACAGTATGCAAGTATTCTTAAGTATTGAAATTAGAATATC 2160  
 ATGAAATAATCAAAACATACAATGCAATGGCAAGTAGGTATGCATATTCAAGAGACTCTC 2220  
 CATTTTGCAAGCTGTGAAGAAAATGTCGTATAAGTTGCTATAAGTTGGGTAGATTCTT 2280  
 GAGAAGCATTCTCATATAATTCACTGAAAGAACCTGATAAATTGACCCACTGTAACTTA 2340  
 GCCACTGATGAAACCTTAAAGCTGAGTATTATAACACCTGATTGTGATTCTTATATAT 2400  
 TCAAAATGCAATTGGTATTGGCTGGCCCTGCTCCCCATCTCTCTTGCCTCATAGATT 2460  
 AGCTATGTTGGGAAAGCACATGCTGCTCTAGGAATATCTCCAAATAAAGCTGTAACTATT 2520  
 TGGTGGAAAAA<sup>n</sup>  
 13

FIG. 6-1

... Exon1	... Exon1	MELPARGWWILCAAAALIVACARG---DPA <u>S</u> KSP <u>S</u> CCGEV <u>R</u> O <u>I</u> YAKGFS-L MAR <u>F</u> GLP <u>A</u> L <u>C</u> T <u>L</u> AV <u>S</u> AA <u>L</u> A <u>A</u> ----EL <u>K</u> SK <u>S</u> <u>C</u> EV <u>R</u> RL <u>V</u> SK <u>G</u> FN-K MP <u>S</u> WIGAV <u>I</u> UPL <u>L</u> LL <u>S</u> P <u>G</u> A-----DV <u>K</u> AR <u>S</u> <u>C</u> EV <u>R</u> QAY <u>G</u> AKGFS-L	46 44 44
... Exon1	... Exon1	MAGTVRTACLVAM <u>L</u> SLDFPGQAQP <u>PPPP</u> DATCH <u>Q</u> VR <u>SFF</u> Q-R <u>L</u> Q <u>P</u> G <u>L</u> MDAQ <u>T</u> WPVGFR <u>C</u> LL <u>L</u> ALVG <u>S</u> <u>A</u> S-----EGVQT <u>C</u> EVR <u>K</u> L <u>F</u> QWR <u>L</u> LG <u>A</u> V	49 45
	Exon2>	SDVP <u>QAEI</u> SGE <u>H</u> IRIC-P <u>QGYT</u> C <u>CT</u> SEMEEN <u>L</u> ANR <u>S</u> HA <u>E</u> LET <u>A</u> LR <u>DSS</u> RV NDAP <u>L</u> HE <u>INGD</u> H <u>L</u> KIC-P <u>QGST</u> <u>C</u> <u>C</u> SQ <u>EMEE</u> E <u>KYS</u> LS <u>OK</u> D <u>DK</u> E <u>K</u> S <u>V</u> V <u>S</u> E <u>QC</u> NH ADIP <u>YQEI</u> AGE <u>H</u> IRIC-P <u>QEYT</u> <u>C</u> <u>C</u> TTEM <u>D</u> <u>K</u> L <u>S</u> Q <u>QSKL</u> E <u>F</u> EN <u>L</u> V <u>E</u> TS <u>H</u> F	95 93 93
	Exon2>	KWVP <u>ETPV</u> PG <u>SDL</u> QV <u>C</u> LP <u>KG</u> P <u>T</u> CC <u>CSR</u> K <u>ME</u> E <u>KY</u> Q <u>L</u> T <u>A</u> R <u>L</u> N <u>M</u> E <u>Q</u> L <u>Q</u> S <u>A</u> S <u>M</u> E RG <u>L</u> D <u>SPRA</u> G <u>PD</u> Q <u>V</u> <u>C</u> <u>IS</u> <u>KK</u> P <u>T</u> <u>C</u> <u>C</u> T <u>R</u> K <u>ME</u> E <u>RY</u> Q <u>IA</u> A <u>R</u> Q <u>D</u> M <u>Q</u> Q <u>F</u> L <u>Q</u> S <u>S</u> ST	99 95
	Exon3>	LO <u>AML</u> T <u>QL</u> R <u>S</u> F <u>DD</u> H <u>F</u> Q <u>H</u> LL <u>D</u> S <u>E</u> R <u>T</u> L <u>Q</u> A <u>T</u> F <u>P</u> G <u>A</u> F <u>G</u> E <u>L</u> Y <u>T</u> Q <u>N</u> A <u>R</u> A <u>F</u> R <u>D</u> LY LO <u>AVF</u> A <u>S</u> R <u>Y</u> K <u>K</u> F <u>DE</u> F <u>E</u> R <u>L</u> E <u>A</u> K <u>S</u> L <u>ND</u> M <u>F</u> V <u>K</u> T <u>Y</u> G <u>H</u> LY <u>M</u> Q <u>N</u> S <u>E</u> LL <u>F</u> K <u>D</u> LF VR <u>T</u> T <u>F</u> V <u>S</u> R <u>H</u> K <u>K</u> F <u>DE</u> F <u>E</u> R <u>L</u> E <u>A</u> K <u>S</u> L <u>ND</u> M <u>F</u> V <u>R</u> T <u>Y</u> G <u>M</u> LY <u>M</u> Q <u>N</u> S <u>E</u> V <u>F</u> Q <u>D</u> LF	145 143 143
	Exon3>	L <u>K</u> E <u>FL</u> I <u>LI</u> O <u>A</u> V <u>V</u> Q <u>E</u> A <u>F</u> E <u>I</u> V <u>V</u> R <u>H</u> A <u>N</u> Y <u>T</u> NAME <u>K</u> NN <u>Y</u> PS <u>L</u> T <u>P</u> Q <u>A</u> F <u>E</u> F <u>V</u> G <u>E</u> F <u>F</u> L <u>K</u> E <u>FL</u> I <u>SI</u> R <u>N</u> AAA <u>A</u> F <u>Q</u> E <u>T</u> L <u>E</u> L <u>I</u> K <u>Q</u> A <u>E</u> N <u>Y</u> NT <u>S</u> I <u>L</u> F <u>C</u> ST <u>Y</u> R <u>M</u> AA <u>A</u> S <u>V</u> Q <u>E</u> F <u>F</u>	149 145
GPC1	GPC1	MELPARGWWILCAAAALIVACARG---DPA <u>S</u> KSP <u>S</u> CCGEV <u>R</u> O <u>I</u> YAKGFS-L MAR <u>F</u> GLP <u>A</u> L <u>C</u> T <u>L</u> AV <u>S</u> AA <u>L</u> A <u>A</u> ----EL <u>K</u> SK <u>S</u> <u>C</u> EV <u>R</u> RL <u>V</u> SK <u>G</u> FN-K MP <u>S</u> WIGAV <u>I</u> UPL <u>L</u> LL <u>S</u> P <u>G</u> A-----DV <u>K</u> AR <u>S</u> <u>C</u> EV <u>R</u> QAY <u>G</u> AKGFS-L	46 44 44
GPC4	GPC4	MAGTVRTACLVAM <u>L</u> SLDFPGQAQP <u>PPPP</u> DATCH <u>Q</u> VR <u>SFF</u> Q-R <u>L</u> Q <u>P</u> G <u>L</u> MDAQ <u>T</u> WPVGFR <u>C</u> LL <u>L</u> ALVG <u>S</u> <u>A</u> S-----EGVQT <u>C</u> EVR <u>K</u> L <u>F</u> QWR <u>L</u> LG <u>A</u> V	49 45
GPC6	GPC6	SDVP <u>QAEI</u> SGE <u>H</u> IRIC-P <u>QGYT</u> C <u>CT</u> SEMEEN <u>L</u> ANR <u>S</u> HA <u>E</u> LET <u>A</u> LR <u>DSS</u> RV NDAP <u>L</u> HE <u>INGD</u> H <u>L</u> KIC-P <u>QGST</u> <u>C</u> <u>C</u> SQ <u>EMEE</u> E <u>KYS</u> LS <u>OK</u> D <u>DK</u> E <u>K</u> S <u>V</u> V <u>S</u> E <u>QC</u> NH ADIP <u>YQEI</u> AGE <u>H</u> IRIC-P <u>QEYT</u> <u>C</u> <u>C</u> TTEM <u>D</u> <u>K</u> L <u>S</u> Q <u>QSKL</u> E <u>F</u> EN <u>L</u> V <u>E</u> TS <u>H</u> F	95 93 93
GPC3	GPC3	KWVP <u>ETPV</u> PG <u>SDL</u> QV <u>C</u> LP <u>KG</u> P <u>T</u> CC <u>CSR</u> K <u>ME</u> E <u>KY</u> Q <u>L</u> T <u>A</u> R <u>L</u> N <u>M</u> E <u>Q</u> L <u>Q</u> S <u>A</u> S <u>M</u> E RG <u>L</u> D <u>SPRA</u> G <u>PD</u> Q <u>V</u> <u>C</u> <u>IS</u> <u>KK</u> P <u>T</u> <u>C</u> <u>C</u> T <u>R</u> K <u>ME</u> E <u>RY</u> Q <u>IA</u> A <u>R</u> Q <u>D</u> M <u>Q</u> Q <u>F</u> L <u>Q</u> S <u>S</u> ST	99 95
GPC5	GPC5	LO <u>AML</u> T <u>QL</u> R <u>S</u> F <u>DD</u> H <u>F</u> Q <u>H</u> LL <u>D</u> S <u>E</u> R <u>T</u> L <u>Q</u> A <u>T</u> F <u>P</u> G <u>A</u> F <u>G</u> E <u>L</u> Y <u>T</u> Q <u>N</u> A <u>R</u> A <u>F</u> R <u>D</u> LY LO <u>AVF</u> A <u>S</u> R <u>Y</u> K <u>K</u> F <u>DE</u> F <u>E</u> R <u>L</u> E <u>A</u> K <u>S</u> L <u>ND</u> M <u>F</u> V <u>K</u> T <u>Y</u> G <u>H</u> LY <u>M</u> Q <u>N</u> S <u>E</u> LL <u>F</u> K <u>D</u> LF VR <u>T</u> T <u>F</u> V <u>S</u> R <u>H</u> K <u>K</u> F <u>DE</u> F <u>E</u> R <u>L</u> E <u>A</u> K <u>S</u> L <u>ND</u> M <u>F</u> V <u>R</u> T <u>Y</u> G <u>M</u> LY <u>M</u> Q <u>N</u> S <u>E</u> V <u>F</u> Q <u>D</u> LF	145 143 143
GPC3	GPC3	L <u>K</u> E <u>FL</u> I <u>LI</u> O <u>A</u> V <u>V</u> Q <u>E</u> A <u>F</u> E <u>I</u> V <u>V</u> R <u>H</u> A <u>N</u> Y <u>T</u> NAME <u>K</u> NN <u>Y</u> PS <u>L</u> T <u>P</u> Q <u>A</u> F <u>E</u> F <u>V</u> G <u>E</u> F <u>F</u> L <u>K</u> E <u>FL</u> I <u>SI</u> R <u>N</u> AAA <u>A</u> F <u>Q</u> E <u>T</u> L <u>E</u> L <u>I</u> K <u>Q</u> A <u>E</u> N <u>Y</u> NT <u>S</u> I <u>L</u> F <u>C</u> ST <u>Y</u> R <u>M</u> AA <u>A</u> S <u>V</u> Q <u>E</u> F <u>F</u>	149 145
GPC5	GPC5	MELPARGWWILCAAAALIVACARG---DPA <u>S</u> KSP <u>S</u> CCGEV <u>R</u> O <u>I</u> YAKGFS-L MAR <u>F</u> GLP <u>A</u> L <u>C</u> T <u>L</u> AV <u>S</u> AA <u>L</u> A <u>A</u> ----EL <u>K</u> SK <u>S</u> <u>C</u> EV <u>R</u> RL <u>V</u> SK <u>G</u> FN-K MP <u>S</u> WIGAV <u>I</u> UPL <u>L</u> LL <u>S</u> P <u>G</u> A-----DV <u>K</u> AR <u>S</u> <u>C</u> EV <u>R</u> QAY <u>G</u> AKGFS-L	46 44 44

## FIG. 6-2

GPC1	SELLRYRGANLHLEETIAEFWARILLERLFKQL-HPQLLLPD-DYLDCLG	193
GPC4	VELKRYYYVGVNVNLEEMINDFWARILLERMF-RIVNSQYHFTD-EYLE <u>C</u> VS	191
GPC6	TELKRYYTGGNVNLEEMINDFWARILLERMF-QLINPQYHFTS-EYLE <u>C</u> VS	191
GPC3	TDVSLIYLGSIDINVDMWNELFDLSLFPVIYTQLMNPGLPDSALDINECL-	198
GPC5	TDVGLYLFADVNPEEFVNRRFFDSLFLPVYNNHLINPGVTDSLEYSE <u>C</u> I-	194
GPC1	KQA---EALRPFGAAPPRLRRAFAARSFVQGLGVASDVVRKVAQ	239
GPC4	KYT----EQLKPFGDVPRKLKLQVTRAFAARTFAQGLAVAGDVVSKSV	237
GPC6	KYT---DQLKPFGDVPRLKLKIQVTRAFAARTFVQGLTVGREVANRVS <u>A</u>	237
GPC3	RGA---RRDLKVFGNEPKLIMTOVSKSLQVTRIFLQALNLGIEVINTTDH	245
GPC5	RMA---RRDVSPPFGNIPQRVMQGMGRSLLPSRTFLQALNLGIEVINTTDY	241
Exon4>		
GPC1	VPLIGPECRAVMKLVYCMAHCLGVPGARPDCPDYCRNVLKGC <span style="background-color: #cccccc;">L</span> ANQADLDAE	289
GPC4	VNPTTAQC <span style="background-color: #cccccc;">T</span> HALLKMLY <span style="background-color: #cccccc;">C</span> SH <span style="background-color: #cccccc;">C</span> RGLVTVKPC <span style="background-color: #cccccc;">C</span> NYCSNI <span style="background-color: #cccccc;">C</span> NMRG <span style="background-color: #cccccc;">C</span> LANQGDLDFE	287
GPC6	VSPTPGC <span style="background-color: #cccccc;">C</span> IRALM <span style="background-color: #cccccc;">M</span> MLY <span style="background-color: #cccccc;">C</span> Y <span style="background-color: #cccccc;">C</span> RGLPTVRP <span style="background-color: #cccccc;">C</span> NNY <span style="background-color: #cccccc;">C</span> LNVMRK <span style="background-color: #cccccc;">C</span> LANQADLDT <u>E</u>	287
GPC3	LKFSSKDCGRMLITRMWYC <span style="background-color: #cccccc;">S</span> YCQGLMMVKPCGGYC <span style="background-color: #cccccc;">C</span> NVVMQGCMAGVVELDKY	295
GPC5	LHF <span style="background-color: #cccccc;">S</span> KE <u>C</u> SRALLKMQY <span style="background-color: #cccccc;">C</span> PH <span style="background-color: #cccccc;">C</span> OGLALT <span style="background-color: #cccccc;">K</span> P <span style="background-color: #cccccc;">C</span> MGY <span style="background-color: #cccccc;">C</span> LNVMRGC <span style="background-color: #cccccc;">C</span> LAHMAELNPH	291
Exon5>		
GPC1	WRNLLDSMVLTDKFWGTSGVESVIGSVHTWIAEAINALQDNRDTLTAKV	339
GPC4	WNNFIDAMIMVAERLEGPFNIESVMDPIDVVKISDAIMNMQDNSVQSQKV	337
Exon6>		

FIG. 6-3

GPC6	WNLFIDAMLLVAERLEGPFNIESVMDPIDVKISEAIMMQENSMQVSAKV ^	337
		Exon4>
GPC3	WREYIISLEELNVGMYRIDMENVLLGLFESTIHDSDIQYVQKNAGKLTTI WHAYIIRSLELLSDAMHGTIDIGHVLLNFHLVNDAVLQAHLNQKLIQEVL ^ ^	345
GPC5		341
		Exon7>
GPC1	IQGGCGNPKVNPQGPGPPEEKRRRG-----KLAPRERPPS--GT-LEKLVS FQGGCGPPKPPLPAGRISRSISESAFSARFRPHPEERPTTAAGTSLSLRLVT FQGGCGQPKPAALARSAAPEN-ENTRFPRYNPEERPTTAAGTSLSLRLVT ^ ^	380
GPC4		387
GPC6		386
		Exon5>
GPC3	GKICIAHSQQRQYRSAYYPEDLFDIKVKVLKVAHVEHEET-----LSSRRR NRICGRPVRTPTQSPRCSEFDQSKEKHGMKTTTRNSEET-----LANRRK ^	389
GPC5		385
		Exon8>
GPC1	EAKAQLRDVQDFEWISLPGTLCS-EKMALSTASDDR-CWNGMARG-RYLPE DVKDALKQAKKFWSSLPSNV <u>C</u> NDERMAAGNGNEDD- <u>C</u> WNGKGKS-RYLFA DIKEKLIKLSKKVWSALPYTIC <u>K</u> DESVTAGTSNEEE- <u>C</u> WNGHSKA-RYLPE ^	427
GPC4		435
GPC6		434
		Exon6>
GPC3	ELIQKIKSFISFYSSALPGYICSHSPVAE-----NDTICWNGQELVERYSQK EFINSRLRYRSFYGGGLADQI <u>C</u> ANELAAA----DGLP <u>C</u> WNGDEDIVKSYTQR ^	435
GPC5		431

21/33

## F|G.6 - 4

VMGDGLANQINNPEVEVDITKPDMTI-----RQQIMQLKIMTNRLRSAYNG 473  
 VTAGNGLIVNOGNNPEVQVDTSKPDILI-----LREQIMALRVMTSKMKNAHYNG 481  
 IMNDGLTNQINNPEVVDITRDPDTFI-----RQQIMALRVMTNKLKNAYNG 480

## Exon 7 &gt;

AARNGMKNQFNHLIELKMKGPEPVVSQIIDKLKHINQOLLRTMSMPKGR----  
 VVGNGNIKAQSGNPEVKVKGIDPVINQIIDKLKHVVQOLLQGRSPKPDK----  
 ^ ^

## Exon 9 &gt;

NDVDFQDASDDGGSGSGSGDGCLDDL--CGRKVSRKSSSSRTP-----  
 NDVDFFDISDESSSGEGSGSGCEYQQ--CPSEFDYNATDH-----  
 NDVNFDQTDSSSGSGSGCMDDV--CPTEFEFVTTEA-----  
 ^

## Exon 8 &gt;

----VLDKNLDEEGFESGD-CGDDDEDECIGGSGDGMIKVKKNQLRFLAEALA 527  
 WELIQLGSGGGMVEQVSGD-C-DDEDGC-GGSGSG--EVKRITLKIDWMP 523  
 ^

----LTHALPGLSEQEGQKTSAASCPQPPT--FLLPLLFLALTVARPRWR 558  
 -----ARKSANEKADSAGVRPGQAQYLTVFCILFLVMORE---WR 556  
 -----PAVDPDRREVDSAAQRGHSILLSLTCI-VLAQRLC---R 555

YDLDDAPGSQQATPKDNELISTFHNLGNVHSPIKLITSMMAISVVCFEFFLVH 580  
 DDMNFSDVQIHQTDGTGSTDITGACAVATES---MTFTLISVVMLLPGIW 572

GPC1  
 GPC4  
 GPC6

GPC3  
 GPC5

GPC3  
 GPC5

GPC1  
 GPC4  
 GPC6

GPC3  
 GPC5

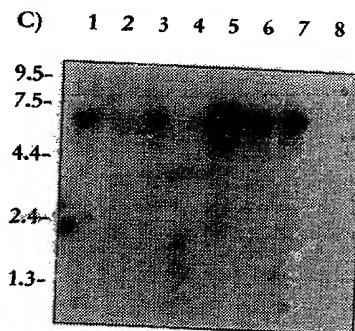
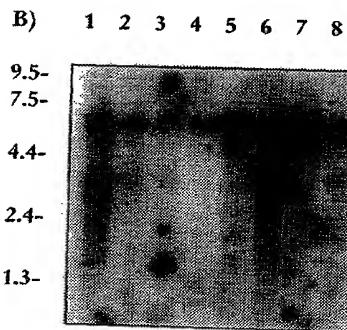
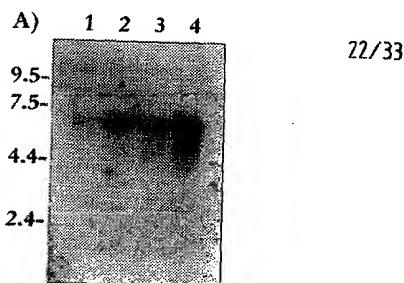
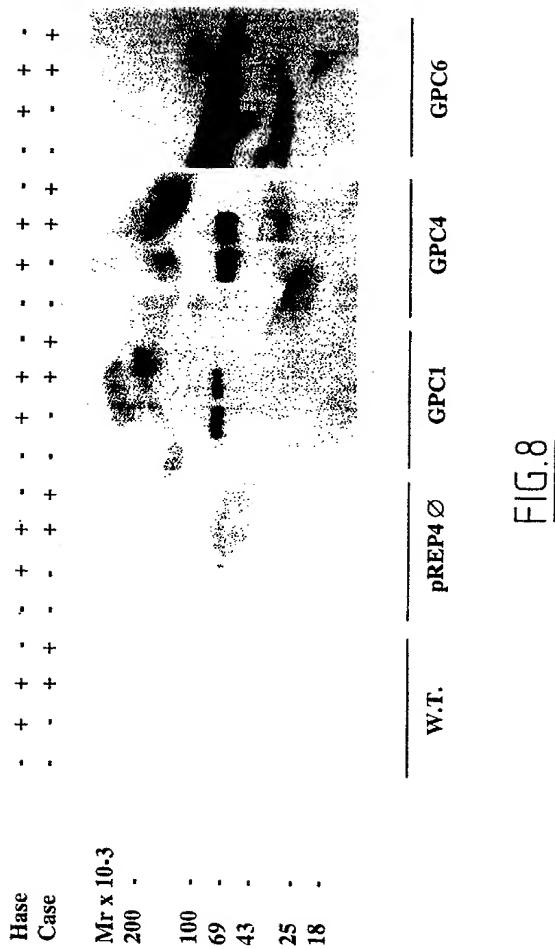


FIG. 7



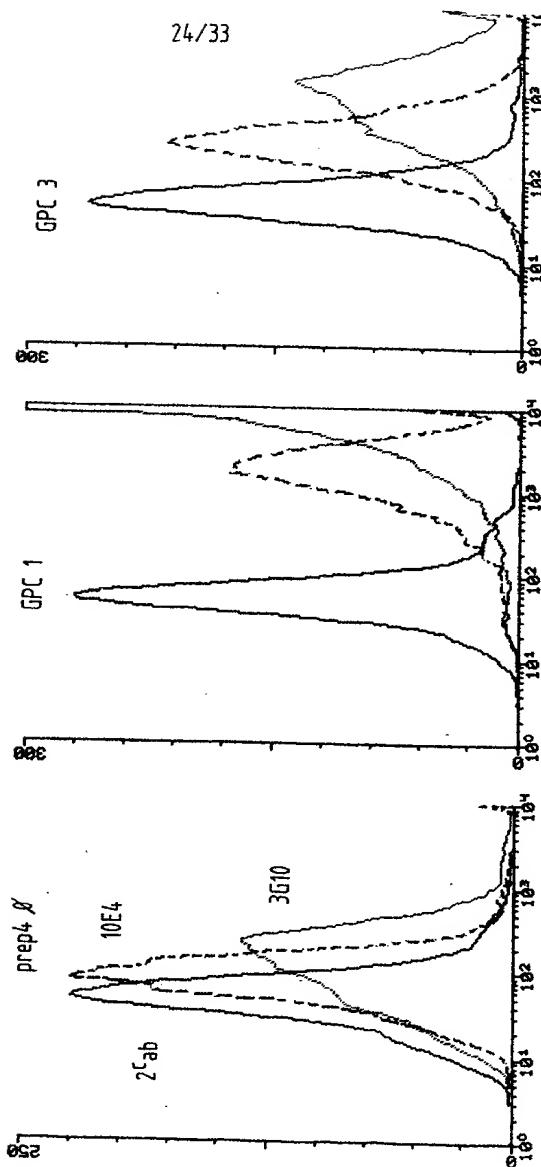


FIG. 9-1

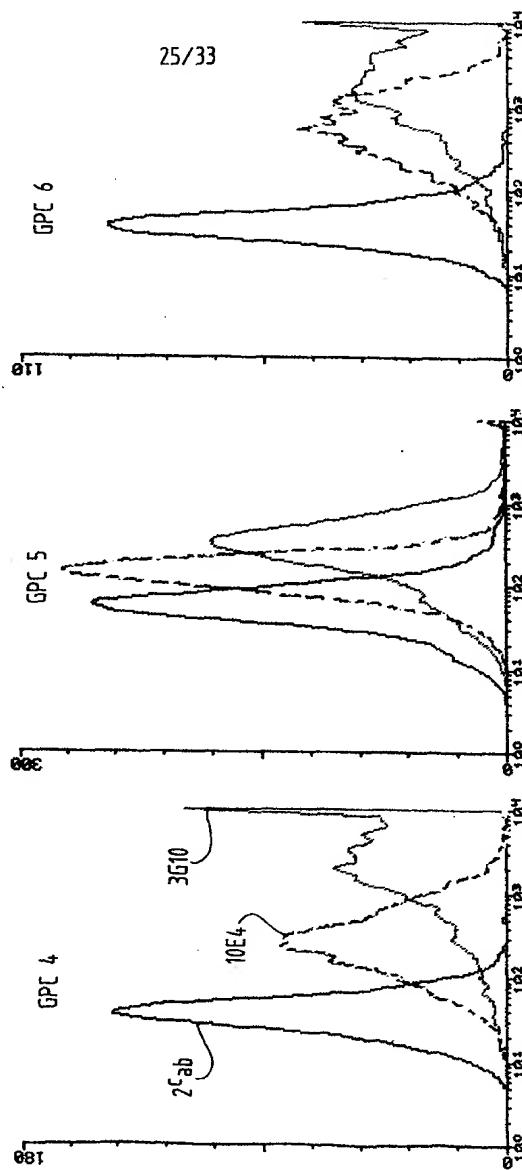


FIG. 9-2

26/33

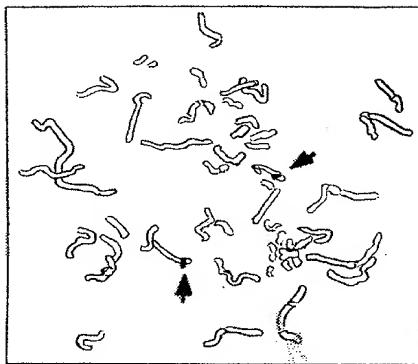


FIG. 10B

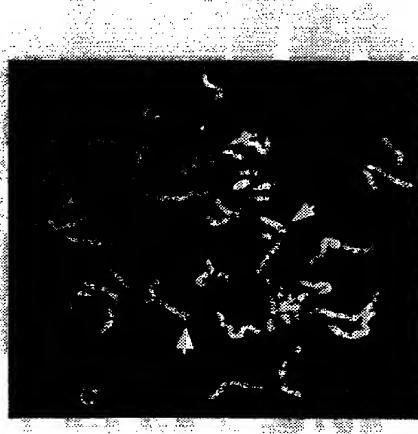


FIG. 10A

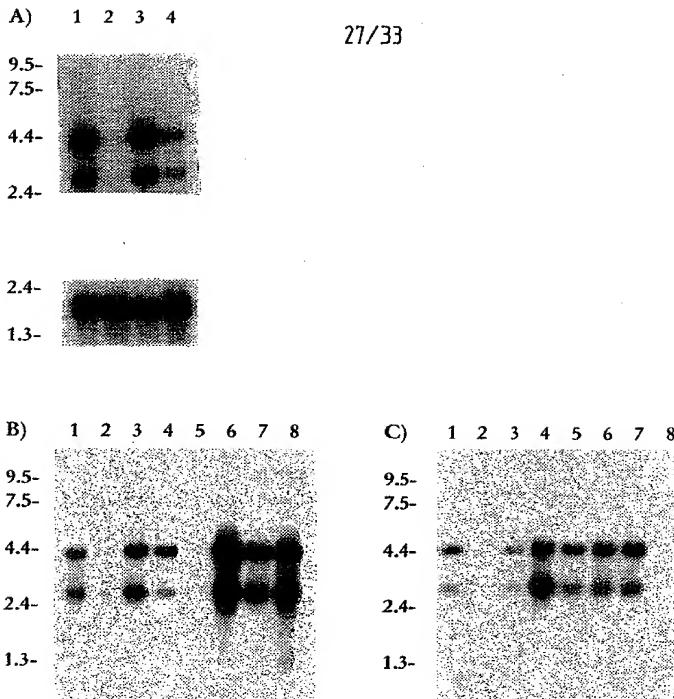


FIG.11

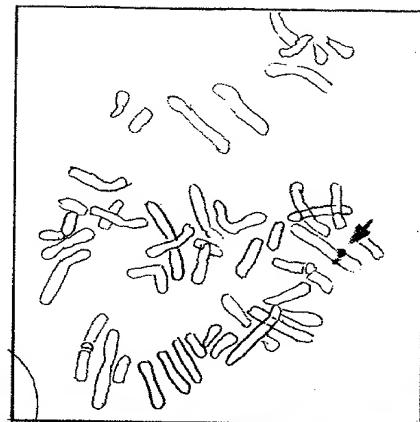


FIG.12A

29/33

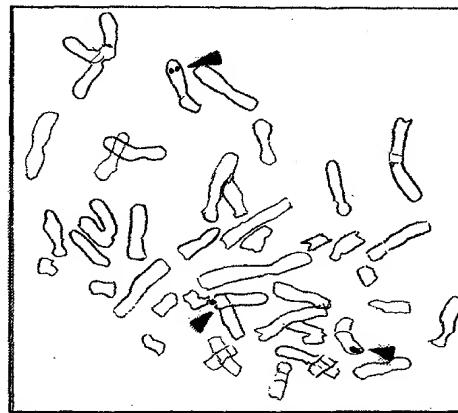


FIG.12B

30/33.

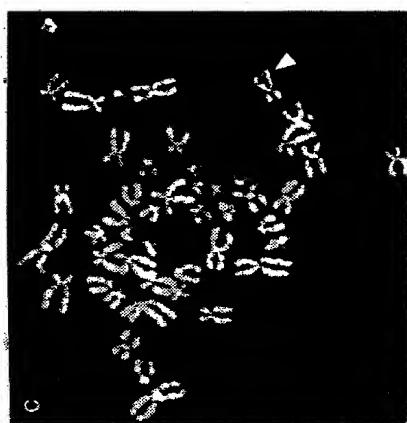
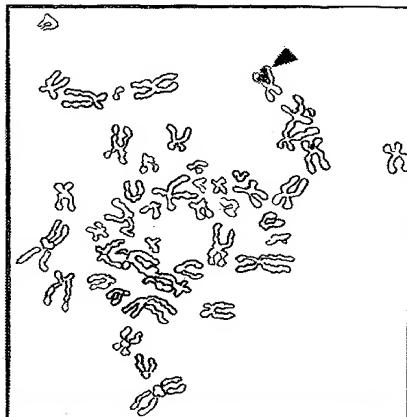


FIG. 12C

31/33

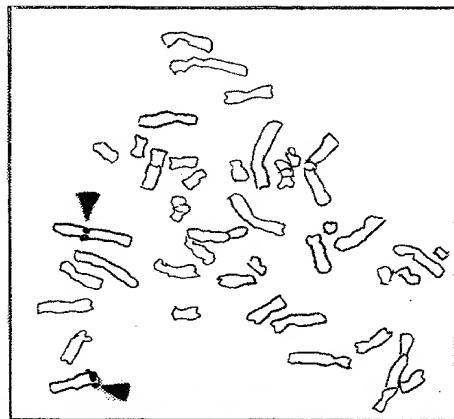


FIG.12D

32/33

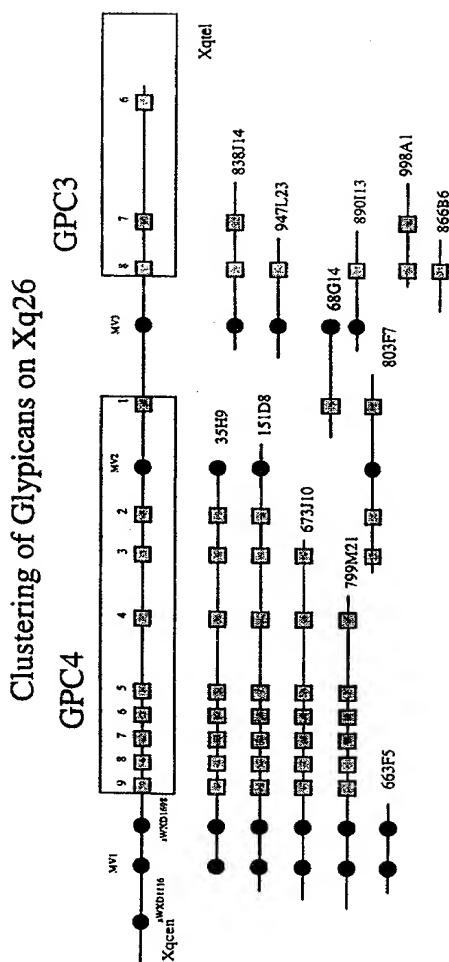


FIG. 13A

FIG.13B) Clustering of Glycans on Xq26

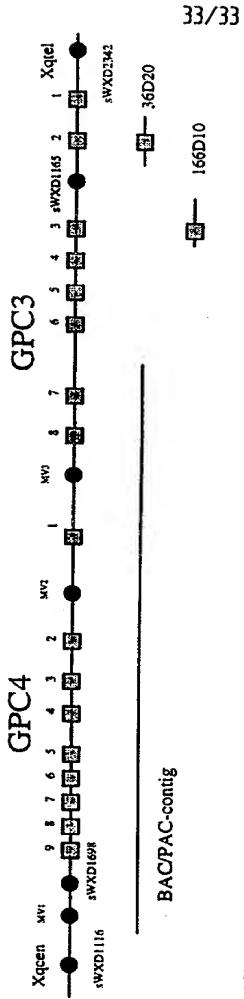


FIG.13C) Glycan deletions in SGBS-patients

